DOCUMENT RESUME

ED 411 215 SP 037 505

AUTHOR	Stronge, James, Ed.
TITLE	Evaluating Teaching: A Guide to Current Thinking and Best Practice.
ISBN	ISBN-0-8039-6356-4
PUB DATE	1997-00-00
NOTE	329p.
AVAILABLE FROM	Corwin Press, Inc., 2455 Teller Road, Thousand Oaks, CA 91320; e-mail: order[at sign]corwin.sagepub.com; fax: 805-499-0871 (softcover: ISBN-0-8039-6356-4, $32.95; hardcover: ISBN-0-8039-6355-6, $69.95).
PUB TYPE	Books (010) -- Collected Works - General (020) -- Guides - Non-Classroom (055)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Classroom Observation Techniques; Elementary Secondary Education; *Evaluation Methods; Faculty Development; *Instructional Effectiveness; Legal Responsibility; Program Development; Program Implementation; Standards; *Teacher Effectiveness; *Teacher Evaluation; Teacher Qualifications; Teachers; Theory Practice Relationship

ABSTRACT

This guide presents current research and thinking about teacher evaluation and combines that research with practice. Chapters contain illustrations and examples to make a research-practice connection and present a comprehensive approach to designing, implementing, and monitoring quality teacher-evaluation systems. Chapters include: (1) "Improving Schools through Teacher Education" (James H. Stronge); (2) "Building the Foundation: Teacher Roles and Responsibilities" (Patricia H. Wheeler and Michael Scriven); (3) "Legal Considerations in Designing Teacher Evaluation Systems" (Pamela D. Tucker and Kay P. Kindred); (4) "Applying the Personnel Evaluation Standards to Teacher Evaluation" (James R. Sanders); (5) "Classroom-Based Assessments of Teaching and Learning" (Chad D. Ellett); (6) "Client Surveys in Teacher Evaluation" (James H. Stronge and Laura P. Ostrander); (7) "Indicators of Student Learning in Teacher Evaluation" (Andrew A. McConney, Mark D. Schalock, and H. Del Schalock); (8) "Portfolios in Teacher Evaluation" (Kenneth Wolf, Gary Lichtenstein, and Cynthia Stevenson); (9) "Teacher Self-Evaluation" (Peter W. Airasian and Arlen Gullickson); (10) "Conducting a Successful Evaluation Conference" (Virginia M. Helm); (11) "Dealing Positively with the Nonproductive Teacher" (Mary Jo McGrath); and (12) "Linking Teacher Evaluation and Professional Development" (Joyce Annunziata). Chapters contain references. (JLS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SP

ED 411 215

Evaluating Teaching

Editor:
James H. Stronge

SP037505

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

G. Alkema

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Evaluating Teaching

Evaluating Teaching

A Guide to Current Thinking and Best Practice

Editor:
James H. Stronge

CORWIN PRESS, INC.
A Sage Publications Company
Thousand Oaks, California

Copyright © 1997 by Corwin Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

For information:

Corwin Press, Inc.
A Sage Publications Company
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@corwin.sagepub.com

SAGE Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

SAGE Publications India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Main entry under title:

Evaluating teaching: A guide to current thinking and best practice / edited by James H. Stronge.
p. cm.
Includes bibliographical references and index.
ISBN 0-8039-6355-6 (cloth: acid-free paper). —
ISBN 0-8039-6356-4 (pbk.: acid-free paper)
1. Teachers—Rating of—United States. 2. Teacher effectiveness—United States. I. Stronge, James H.
LB2838.E847 1997
371.14'4—dc21 96-51310

This book is printed on acid-free paper.

97 98 99 00 01 02 03 10 9 8 7 6 5 4 3 2 1

Editorial Assistant: Kristen L. Green
Production Editor: Diana E. Axelsen
Production Assistant: Denise Santoyo
Typesetter/Designer: Janelle LeMaster
Cover Designer: Marcia R. Finalyson

Contents

Preface xi

About the Authors xiii

1. **Improving Schools Through Teacher Evaluation**
James H. Stronge 1

Why Is There a Need for High-Quality Teacher Assessment and Evaluation? 2

What Are the Basic Purposes of a High-Quality Teacher Assessment and Evaluation System? 3

What Are the Key Features of an Effective Teacher Assessment and Evaluation System? 6

How Can a Teacher Assessment and Evaluation System Be Designed for School Improvement and Personal Growth? 12

Summary 17

PART I: Designing a Teacher Evaluation System

2. **Building the Foundation: Teacher Roles and Responsibilities**
Patricia H. Wheeler and Michael Scriven 27

Purposes of Teacher Evaluation Systems 28

What Are Some Possible Foundations? 29

How Does One Choose a Foundation? 37
How Are Teacher Roles and Responsibilities Determined? 39
What Are the Next Steps in Developing a Teacher Evaluation System? 46
Summary 54

3. **Legal Considerations in Designing Teacher Evaluation Systems**
Pamela D. Tucker and Kay P. Kindred 59
Why Evaluate? 60
What Are the Sources of Legal Requirements in Teacher Evaluation? 63
What Are the Recommended Components of an Evaluation System? 70
What Are the Implementation Aspects of an Evaluation System? 77
What Are the General Principles for Teacher Evaluation? 84
Conclusion 85

4. **Applying the Personnel Evaluation Standards to Teacher Evaluation**
James R. Sanders 91
Background 92
How Were the Personnel Evaluation Standards Developed? 93
What Are the Implications of the Personnel Evaluation Standards for Teacher Evaluation? 95
Suggestions for Using the Standards 101
A Final Word 102

PART II: Assessing Teacher Performance

5. **Classroom-Based Assessments of Teaching and Learning**
Chad D. Ellett 107
A Brief History of Classroom-Based Assessment of Teachers and Teaching 108

What Is the Problem With Traditional Classroom-Based Assessment? 113
How Can Classroom-Based Assessment Be More Productive? 114
Some Suggestions for Solutions 115
The System for Teaching and Learning Assessment and Review (STAR) 118
A Call for Multiple Methodologies 121
Synthesis and Projection 124

6. **Client Surveys in Teacher Evaluation**
James H. Stronge and Laura P. Ostrander 129
Why Should Client Data Be Used in Teacher Evaluation? 130
What Are the Benefits of Including Parents in the Teacher Evaluation Process? 135
Can Students Provide Reliable and Useful Measures of Teaching Effectiveness? 144
How Can Feedback From Peers Be Used Most Effectively in Teacher Evaluation? 150
Using Client Surveys as Part of Multiple Data Sources in Teacher Evaluation 156
Summary 157

7. **Indicators of Student Learning in Teacher Evaluation**
Andrew A. McConney, Mark D. Schalock, and H. Del Schalock 162
Past Efforts and Their Consequences 163
Which Emerging Systems Are Promising? 166
What Are Current Approaches to Assessing Learning and Their Evaluative Potential? 173
What Is the Proper Balance in Using Student Learning in the Evaluation of Teachers? 183
Summary 186

8. **Portfolios in Teacher Evaluation**
Kenneth Wolf, Gary Lichtenstein, and Cynthia Stevenson 193
What Is a Teaching Portfolio? 194
What Purposes Might a Teaching Portfolio Serve? 195

What Might Be Included in a Teaching Portfolio? 196
How Might the Portfolio Process Unfold? 197
How Should a Teaching Portfolio Be Evaluated? 198
Examples From Teaching Portfolios 204
Putting Portfolios Into Practice 209
Conclusion 213

9. **Teacher Self-Evaluation**
Peter W. Airasian and Arlen Gullickson 215
Why Is Teacher Self-Evaluation Important? 216
What Is the Role of Reflection in Teacher Self-Evaluation? 219
Getting Started With Self-Evaluation 228
Strategies for Teacher Self-Evaluation 229
What Are the Characteristics for Successful Self-Evaluation? 236
Building a Support Structure for Self-Evaluation 239
Conclusion 241

PART III: Implementing the Teacher Evaluation System

10. **Conducting a Successful Evaluation Conference**
Virginia M. Helm 251
What Are the Purposes of the Evaluation Conference? 252
Preparing for the Evaluation Conference 253
Conducting the Evaluation Conference 256
What Are the Characteristics of Effective Evaluation Conferences? 258
How Can Principals Facilitate Productive Teacher Talk? 260
Concluding the Interview 265
Summary 265

11. **Dealing Positively With the Nonproductive Teacher**
Mary Jo McGrath 270
How Do We Make Accountability an Opportunity, Not a Punishment? 271

How Does the Law of Accountability Consider the Whole? 272
What Is the Architecture of Accountability? 280
Evaluation and Documentation: Why Are They So Important? 282
Understanding the Culture and Ethics of Accountability 283
Conclusion 285

12. Linking Teacher Evaluation and Professional Development
Joyce Annunziata 288
What the Literature Reveals About Teacher Evaluation Practices 289
How Is Professional Development Viewed Today? 291
What Is New About Professional Development? 291
Professional Development: The Missing Link Between Evaluation and Improved Performance 293
When Does Professional Development in an Evaluation System Stop? 294
Symbolic Training 295
Who Should Deliver the Professional Development Program for Administrators, Teachers, and District Staff? 297
What Is Effective Professional Development for Implementation of the Evaluation System? 298
Professional Development Is the Linchpin Between Performance and Growth for the Administrator and Teacher Practitioner 298
Summary 299

Author Index 303

Subject Index 311

Preface

There is renewed interest in the role of teacher evaluation as a fundamental aspect of school improvement. To a large extent, this interest comes from the realization that any significant improvement in schooling must have the teacher at its heart. And, just as there is a rational connection between school improvement and teacher performance, there is a necessary and rational connection between teacher improvement and teacher evaluation.

This volume synthesizes current research and thinking about teacher evaluation and blends that research with practice. Each chapter is rich with illustrations and examples as we attempt to make a strong research-practice connection. In addition, a question-and-answer format is employed in each chapter to enhance the utility of the book and its practical applications.

The book is organized into 12 chapters: an introductory chapter and three major sections composed of 11 additional chapters. In the introductory chapter, I suggest that "a conceptually sound and properly implemented evaluation system for teachers (indeed, for all educators) is a vital component of any successful reform effort." I then discuss key features of effective teacher evaluation systems and offer one model for designing a quality teacher evaluation system for school improvement and teacher growth.

Part I, Designing a Teacher Evaluation System, begins with a chapter by Patricia Wheeler and Michael Scriven in which they describe foundations for teacher evaluation and then focus on roles and responsibilities of the teacher as a desirable foundation. In Chapter 3, Pamela Tucker and Kay Kindred discuss the legal context and parameters for designing and conducting teacher evaluations. In Chapter 4, James Sanders provides background and implications for using the Personnel Evaluation Standards in teacher evaluation.

Part II, Assessing Teacher Performance, provides a rationale and explores application issues for using multiple data sources in assessing teacher performance. In Chapter 5, Chad Ellett offers a call for going beyond teacher evaluation by using classroom-based observation for assessing teaching and learning. Laura Ostrander and I offer guidance for using parent, student, and peer feedback in Chapter 6. In Chapter 7, Andrew McConney, Mark Schalock, and Del Schalock discuss promising approaches to using student learning in the evaluation of teachers' work. Kenneth Wolf, Gary Lichtenstein, and Cynthia Stevenson describe the benefits, liabilities, and applications of teacher portfolios in Chapter 8. In the last chapter in Part II, Peter Airasian and Arlen Gullickson describe a process in which teachers make judgments about their own performance for the purpose of self- improvement.

Part III, Implementing the Teacher Evaluation System, explores often neglected but important issues in the successful application of an effective evaluation system. In Chapter 10, Virginia Helm draws heavily from the business and professional literature to describe how evaluation conferences can serve as a key component in an ongoing communication and feedback cycle. In Chapter 11, Mary Jo McGrath discusses evaluation factors that should be incorporated in practice on a day-to-day basis. In the final chapter, Joyce Annunziata makes a case for the vital connection between teacher evaluation and professional development in the successful school improvement process.

Taken collectively, the contributions to this volume represent an attempt to provide a coherent and comprehensive approach to the design, implementation, and monitoring of quality teacher evaluation systems. Our goal is to provide a framework for improvement—improvement of the teacher, improvement of the school, and, ultimately, improvement in student performance. My contributors and I trust that our intended readers—school practitioners (principals, staff development specialists, superintendents), graduate students, and researchers interested in teacher evaluation—find the text to be true to this purpose and of value in their work.

—James H. Stronge

About the Authors

Peter W. Airasian is Professor of Education and Chair of the Department of Counseling, Development Psychology, and Research Methods at Boston College. He received his Ph.D. from the University of Chicago, with a specialization in assessment, testing, and evaluation. He is a former high school chemistry and biology teacher. His research focuses on classroom assessment and teacher evaluation. He is the author of *Classroom Assessment* and a member of the founding board of CREATE, the Consortium for Research on Educational Accountability and Teacher Evaluation.

Joyce Annunziata is Senior Executive of the Office of Professional Standards, Dade County Public Schools, Miami, Florida, where she monitors employee performance and serves as an expert witness in litigation in personnel procedures, performance appraisal, and code of ethics. She has a doctorate in Educational Leadership from the University of Miami. Active in the work of the Center for Research on Educational Accountability and Teacher Evaluation (CREATE), she has presented at four National Evaluation Institutes and has reviewed texts and articles from CREATE. She currently serves as a member of the founding board of the CREATE Consortium.

Chad D. Ellett is Professor of Educational Administration and Educational Research Methodology at Louisiana State University, Baton Rouge. He has a long history of research and development work in teacher and personnel evaluation in education and has written extensively in the field. He has developed a variety of assessment and evaluation measures and procedures, and has served as a consul-tant on personnel evaluation to many states, local school districts, and colleges and universities. He currently serves as editor in chief of the *Journal of Personnel Evaluation in Education.*

Arlen Gullickson is Professor of Education and Chief of Staff of the Evaluation Center at Western Michigan University. He is a member of the American Educational Research Association, the National Council on Measurement in Education, the National Education Association, and the Center for Research on Educational Accountability and Teacher Evaluation (CREATE). Currently, he is President of CREATE and Vice Chair of the Joint Committee on Standards for Educational Evaluation.

Virginia M. Helm is Assistant Provost and Professor of Educational Administration and Supervision at Western Illinois University. Her research has focused on the legal aspects of evaluating educators. She has served as a consultant for the Professional Support Personnel Project of the Center for Research on Educational Accountability and Teacher Evaluation. She has recently coauthored a text and a number of articles and technical reports in the area of personnel evaluation. Before she received her doctorate from the University of Iowa, she was a high school English teacher.

Kay P. Kindred is Assistant Professor of Law at the College of William and Mary, Marshall-Wythe School of Law, and a member of the adjunct faculty of the college's School of Education. From 1989 to 1996, she also held the position of Deputy Director of the Institute of Bill of Rights Law, and in that capacity she served as the institute's principal administrator. She received her A.B. from Duke University and her J.D. from Columbia University Law School. She teaches courses in education law, family law, children's rights, and civil rights. She is the author of articles in recent issues of the *Arizona Law Review* and the *Ohio State Law Journal* and is a contributing author to two recently published books, *Corporate Misconduct* (1995) and *A Year in the Life of the Supreme Court* (1995), the latter of which won the 1996 ABA Silver Gavel Award.

Gary Lichtenstein currently is an independent evaluation consultant. He has been director of teacher education programs at Stanford University and the University of Denver. He has researched portfolio development as a teacher educator and has been a consultant to teachers and state officials in California and Colorado. His articles have focused on teacher professional development, including teacher empowerment and professional development. He is currently completing a doctorate in educational administration and policy analysis at Stanford University.

Andrew A. McConney is Associate Professor in the Teaching Research Division, Western Oregon State College. He is engaged in research and development on the assessment and evaluation of pre- and in-service teaching and teacher education, with the aim of enhancing teachers' effectiveness in promoting student learning in the context of Oregon's standards-based design for schools and schooling. Prior to his recent appointment at Western Oregon, he directed the Cross-Cutting Theory Project of the Center for Research on Educational Accountability and Teacher Evaluation at Western Michigan University's Evaluation Center.

Mary Jo McGrath has been a practicing attorney for more than 19 years, focusing on employee performance issues and legal mandates in the workplace. In 1989 she founded her own law office in Santa Barbara, California, specializing in education and personnel law. She is a consultant to the Office of the California Governor on teacher tenure reform and has been the featured legal expert on ABC's *20/20* on the issue of teacher quality and performance. She has authored a staff development program titled the McGrath Template© Communication System for legally fit and ethically principled management. She received her J.D. from Loyola University School of Law.

Laura P. Ostrander currently is an Assessment Specialist with the Virginia Beach City School Division in Virginia. She has more than three decades' experience as a science teacher. In addition, she was instrumental in developing and implementing a client survey program for the Virginia Beach City Public Schools. She received her doctorate in curriculum and instruction from the University of Virginia.

James R. Sanders is Professor of Education at the Evaluation Center at Western Michigan University. He received his Ph.D. in educational research and evaluation from the University of Colorado. He is author or coauthor of numerous articles, monographs, books, and technical reports in the area of education evaluation. He serves as chairman of the Joint Committee on Standards for Educational Evaluation, a coalition of 16 professional organizations concerned with the quality of evaluations done in education.

H. Del Schalock is a Professor in the Teaching Research Division, Western Oregon State College. He has been actively engaged in educational research, development, and improvement in Oregon for more than 30 years. His current professional interests span teacher preparation and licensure for standards-based schools, the redesign of student and teacher work in standards-based schools, and grounded theories of teacher effectiveness—and theories of teaching and learning generally—within the context of standards-based schools.

Mark D. Schalock is Assistant Research Professor in the Teaching Research Division, Western Oregon State College. He has been involved in educational research, development, and evaluation activities at the local, state, and national levels for the past 14 years. His previous work has focused on both individual and institutional improvement and accountability, based on measures of impact. Currently, he codirects a project that aims to connect teacher work meaningfully to student learning through applied performance assessment methodology.

Michael Scriven is Director of the Evaluation and Development Group. He directed the Teacher Evaluation Models Project for 4 years for the federally funded Center for Research on Educational Accountability and Teacher Evaluation at Western Michigan University. He is a former president of the American Educational Research Association and has taught at several universities in the United States and Australia. He has degrees in mathematics and philosophy from the University of Melbourne and a doctorate in philosophy from Oxford University.

Cynthia Stevenson has served as both an elementary and middle-level principal and is currently supervising schools and principals in Jefferson County Schools in Colorado. She also teaches in the Graduate School of Education at the University of Colorado at Denver as an Adjunct Professor in leadership development. In 1992, Colorado Governor Roy Romer appointed her to the State Principal and Administrator Professional Standards Board. During this appointment, she served as chair and worked to implement the Licensure Laws for Educators passed by the Colorado State Legislature. Her doctorate in educational administration and curriculum development is from the University of Colorado at Boulder.

James H. Stronge is Heritage Professor and Coordinator, Educational Policy, Planning, and Leadership Area, at the College of William and Mary in Williamsburg, Virginia. One of his primary research interests is in personnel evaluation, and he has worked with numerous educational organizations to design and develop evaluation systems for teachers and professional support personnel. He is the author or coauthor of numerous articles, books, and technical reports in the area of personnel evaluation, and he served as director of the Evaluating Professional Support Personnel project conducted by the Center for Research on Educational Accountability and Teacher Evaluation (CREATE). Currently, he is a member of the CREATE board and Associate Editor of the *Journal of Personnel Evaluation in Education*. His doctorate in the area of educational administration and planning was received from the University of Alabama.

Pamela D. Tucker is Visiting Assistant Professor of Educational Leadership at the College of William and Mary. She teaches courses in personnel administration, supervision, educational law, and case studies in educational leadership. From 1992 to 1995, she served as Project Coordinator for the Professional Support Personnel CREATE Project (OERI, U.S. Department of Education funded). She is coauthor of the *Evaluation Handbook for Professional Support Personnel* (1995) and has published in professional journals and presented at numerous national conferences on personnel evaluation issues. She has been a classroom teacher and school administrator in a special education setting. She received her doctorate from the College of William and Mary.

Patricia H. Wheeler is the President and Principal Research Associate of EREAPA Associates. She was a Consultant and Senior Research Associate with the federally funded Center for Research on Educational Accountability and Teacher Evaluation at Western Michigan University for 5 years, and a researcher, evaluator, and program administrator at Educational Testing Service for more than 20 years. She has a bachelor's degree in psychology and a master's in counseling from the University of Rochester, an M.B.A. in management from Armstrong College, and a Ph.D. in education from the University of California at Berkeley.

Kenneth Wolf is Assistant Professor of Education at the University of Colorado at Denver, where he is a faculty member in the Research and Evaluation Program. He teaches courses in research methods and assessment for teachers, counselors, and school administrators. He has worked on a variety of projects associated with teacher portfolios, including with the National Board for Professional Teaching Standards and with state departments of education in Colorado and Texas. He was an elementary, middle school, and high school teacher in the United States and in international schools abroad for 10 years, including 2 years as a Peace Corps volunteer in North Africa. His doctorate in curriculum and teacher education is from Stanford University.

1

Improving Schools Through Teacher Evaluation

JAMES H. STRONGE

A conceptually sound and properly implemented evaluation system for teachers is a vital component of an effective school. Regardless of how well an educational program may be designed, the program is only as effective as the people who implement and support it (Stronge, 1993). Despite the fact that proper assessment and evaluation of teachers is fundamental to successful schools and schooling,[1] this part of the personnel process is too frequently neglected—due not to the lack of teacher evaluation but rather to the implementation of poor evaluation systems and poor evaluation practices.

The basic needs in a quality teacher evaluation system are for a fair and effective evaluation based on performance and designed to encourage improvement in both the teacher being evaluated and the school. The purpose of this book is to explore key elements in the construction and implementation of fair and effective teacher

---

AUTHOR'S NOTE: Portions of this chapter are derived from my article "Balancing Individual and Institutional Goals in Educational Personnel Evaluation: A Conceptual Framework," which appeared in *Studies in Educational Evaluation* (1995).

evaluation systems. In this introductory chapter I attempt to set the stage by discussing critical components of a high-quality teacher evaluation system and by identifying how effective teacher evaluation contributes to effective schools. Specifically, I address the following questions:

1. Why is there a need for high-quality teacher assessment and evaluation?
2. What are the basic purposes of a high-quality teacher assessment and evaluation system?
3. What are the key features of an effective teacher assessment and evaluation system?
4. How can a teacher assessment and evaluation system be designed for school improvement and personal growth?

Why Is There a Need for High-Quality Teacher Assessment and Evaluation?

Too often, educational reform has produced disappointing results (Clark & Astuto, 1994) or outright failure (Pogrow, 1996). Fullan (1996) notes that one of the reasons for systemic failure of systemic reforms is fragmentation: "Fragmentation occurs when the pressures—and even the opportunities—for reform work at cross purposes or seem disjointed and incoherent" (p. 420).

An example of working at cross-purposes, of disjointed and incoherent reform, that is played out in schools on a regular basis is the following sequence: (a) Change school policy for a given innovative teacher program, (b) provide some level of staff development on the prospective innovation, (c) ostensibly implement the innovative practice, and (d) continue to use existing evaluation practices. When a reform effort is disconnected from assessment, there is no way to measure the effort's level of success. Such a disconnect is a formula for failure.

A conceptually sound and properly implemented evaluation system for teachers (indeed, for all educators) is a vital component of any successful reform effort. "A rational relationship exists between personnel and programs: If program effectiveness is important and if personnel are necessary for effective programming, then a conceptually sound and properly implemented evaluation system for . . . education personnel is essential" (Stronge, 1993, p. 445).

A dynamic relationship between the teacher and the school exists in a healthy organization: What's good for the organization must also be good for the teacher. This type of synergistic relationship enhances the ability of both the teacher and the school to achieve desired goals. Moreover, balancing individual needs with institutional expectations is essential for fostering productive work environments (March & Simon, 1967, 1993).

If we are correct in assuming that individual and institutional goals are intertwined, then it is logical to consider teacher evaluation as a vehicle to facilitate and assess success for both the teacher (e.g., personal growth and performance improvement) and the school (e.g., goal accomplishment, accountability). Thus, teacher evaluation can and should be considered a vital part of the total improvement/restructuring efforts in education.

Improvement can take numerous forms, including the following:

- Improvement in performance of individual teachers and other educators (administrators, support personnel)
- Improvement of programs and services to students, parents, and community
- Improvement of the school's ability to accomplish its mission

Fostering improvement in teacher evaluation systems means balancing individual and institutional demands. Little (1993) asserts that "the language of reform underestimates the intricate ways in which individual and institutional lives are interwoven" (p. 147). As Fullan (1991) notes, "Combining individual and institutional development has its tensions, but the message . . . should be abundantly clear. You cannot have one without the other" (p. 349).

What Are the Basic Purposes of a High-Quality Teacher Assessment and Evaluation System?

The two most frequently cited purposes of personnel evaluation are accountability and performance improvement (see, for example, Duke, 1990; McLaughlin & Pfeifer, 1988). The accountability purpose reflects the need to determine the competence of teachers in order to ensure that services delivered are safe and effective (McGaghie, 1991), and typically has been viewed as summative in nature. The performance improvement purpose reflects the need for professional growth and development of the individual teacher, and typically has been considered to be formative in nature.[2]

There is room in teacher evaluation systems for both accountability and performance improvement purposes. In fact, evaluation systems that include both accountability and personal growth dimensions are both desirable and necessary if evaluation is to serve the needs of individual teachers and the school and community at large.

> Performance improvement and accountability purposes are not competing, but supportive interests—dual interests that are essential for improvement of educational service delivery. These two roles are inextricably intertwined in the total evaluation process. Moreover, a conceptual framework for [teacher] evaluation should emphasize the dynamic relationship between individual and institution where the needs and interests of one fuse with and support the other. (Stronge, 1995, pp. 131-132)

For teacher evaluation systems to serve dual purposes, however, there must be a rational link between the purposes (Stronge, 1995). McGreal (1988) has argued that multiple purposes of evaluation can be met successfully with a single evaluation system when the system is viewed as one component of a larger mission—furthering the goals of the school. This conception of teacher evaluation ties evaluation not only to teacher improvement but also to school improvement. Thus, a comprehensive teacher evaluation system should be rooted in two broad purposes (Stronge, Helm, & Tucker, 1995):

- It should be *outcome oriented*, contributing to the personal goals of the teacher and to the mission of the program, the school, and the total educational organization, and should provide a fair measure of accountability of performance (i.e., summative focus).
- It should be *improvement oriented*, contributing to the personal and professional development needs of the individual (teacher) as well as improvement within the school (i.e., formative focus).

Outcome Orientation

Unity of purpose is a hallmark of an effective school (Stronge, 1993). Indeed, school- or systemwide purposes should form the basis for all organizational action. Concerning tying organizational purposes to evaluation, Castetter (1981) notes that "a performance appraisal system has its genesis in the broad purposes of the organization" (p. 239). "The importance of establishing school and districtwide goals and having the total personnel function, including teacher evaluation, revolve around the goals" should be obvious in a sound evaluation system (Stronge & Helm, 1991, p. 78). The evaluation system should facilitate not only accomplishment of the school's goals but also compatibility with and support for individual teacher goals. Additionally, if goal accomplishment (for both school and teacher) is fundamental to success, then the evaluation system should reflect this orientation (Stronge & Helm, 1992).

Ralph Tyler (1942) has addressed the outcome/goal orientation that should serve as a basis for teacher evaluation systems:

> A . . . basic assumption involved in evaluation is that the kinds of changes in behavior patterns in human beings which the school . . . seeks to bring about are its educational objectives. The aims of any educational program cannot be stated in terms of the content of the program, or in terms of the methods and procedures followed by the teachers, for these are only means to other ends. Fundamentally, the purposes of education represent these changes in human beings which we hope to bring about through education. (p. 495)

Improvement Orientation

Although a teacher assessment and evaluation system should be oriented toward accomplishing the school's goals, it should also be focused on improvement. Goals typically reflect desired states of being, not existing states. Therefore, if established goals (for both the individual teacher and the school) are to be achieved, an emphasis on improvement and monitoring of progress toward goal accomplishment is inherent in a sound evaluation system (Stronge & Helm, 1992). Stufflebeam (1983) describes this improvement emphasis in evaluation well when he notes that the purpose of evaluation is "not to prove but to improve" (p. 117).

A teacher performance assessment and evaluation system, properly designed and implemented, supports a balanced relationship between school- or districtwide goals and individual teacher professional growth and improvement. For teacher evaluation systems to support professional growth, both organizational barriers (e.g., incompatibility of individual and institutional needs) and personal barriers (e.g., disillusionment, distrust, stress, fear of failure) must be removed (Duke, 1993). Despite the complexities of addressing both improvement and outcome concerns, such a combination is needed. As Saphier (n.d.) states,

> There are those who say supervision must be separated from evaluation because it is impossible for teachers to open up and have productive, growth-oriented dialog with one who judges them. In other words, teacher evaluation is incompatible with stimulating teachers' thinking and growth. We reject that notion. The problem is not that evaluators can't supervise, it is that they cannot supervise often enough. (p. 50)

What Are the Key Features of an Effective Teacher Assessment and Evaluation System?

To achieve a high-quality teacher assessment and evaluation system built upon a dynamic balance between school improvement and teacher improvement, several key features are essential. Among these important concepts and criteria are mutually beneficial goals, emphasis on systematic communication, climate for evaluation, technically sound evaluation systems, and use of multiple data sources. Each of these features is briefly explored in turn on the following pages.

Mutually Beneficial Goals

As described in the previous section, goals that are valued by both the individual teacher and the school are vital to successful teacher evaluation systems. Unless the individual and institutional purposes and goals are mutually beneficial, the efforts of the school are likely to be futile, and teacher evaluations based on those efforts will be meaningless. Scriven (1972) describes the essence of this position: "It is obvious that if the goals aren't worth achieving then it is uninteresting how well they are achieved" (pp. 126-127). Mutually beneficial and supportive goals make the evaluation process and outcomes more acceptable and beneficial to the school community and the teacher alike. Goals that are mutually beneficial for (i.e., compatible with) the individual and the institution are essential. Indeed, if goal accomplishment (for both the school and the teacher) is fundamental to success, then the evaluation system should reflect this balanced perspective (Stronge & Helm, 1992).

Olsen and Bennett (1980) describe performance evaluation as a partnership between organization and employees—a social process that results in a participatory approach to evaluation. It is this mutually beneficial relationship in the workplace that should guide the development and implementation of a relevant, mutually beneficial teacher evaluation system.

Emphasis on Systematic Communication

Teacher evaluation systems should reflect the importance of effective communication in every aspect of the evaluation process, including aspects that are public in nature (e.g., public relations) as well as those that require more private communication (e.g., interpersonal relations) (Stronge, 1995).

One aspect of an emphasis on systematic communication in evaluation is that of public disclosure of those elements about which teachers, administrators, and the general public have the right to be informed. Vital elements for public disclosure in teacher evaluation include the following:

- Establishment of institutional goals
- Determination of evaluation purposes in relation to those goals
- Development of teacher job descriptions that specify roles and responsibilities
- Identification of acceptable standards of performance
- Delineation of procedural guidelines and safeguards embedded in the evaluation system
- Description of the evaluation timeline

From the beginning of discussion regarding the identification of the needs/goals of the school through summative evaluations of performance, stakeholders should know as much about the evaluation system as possible.

For teachers and other educators, certain aspects of public communication and disclosure regarding teacher evaluation are codified in law (e.g., in state statutory requirements for substantive and procedural due process in evaluation decisions). For the general public, this right to know about the school's evaluation system is reflected both in general public policy and in law. For instance, state open meetings acts typically provide for public disclosure and opportunities for public discussion on all relevant issues of policy and practice being considered by local school boards, with a few narrowly defined exceptions (e.g., personal and confidential information about individuals).

Contrasting with the openness in communication described above, another vital aspect of effective communication is more personal and private in nature—that of ongoing two-way communication between the administrator-evaluator and teacher-evaluatee. Good communication between the evaluator and the evaluatee is advantageous for several reasons:

- It allows for the cooperative development of an evaluation plan.
- It provides systematic opportunities for individual skill enhancement and improved performance.
- It provides the teacher with enhanced self-expectations.
- It increases the likelihood that needed changes in performance will take place.
- It allows identification of ways to reach higher standards and correct significant discrepancies.
- It establishes a system of checks and balances for the evaluation process.

Systematic communication between the evaluator and the evaluatee throughout an evaluation cycle minimizes unintended consequences and maximizes organizationally relevant improvement and performance (Cummings & Schwab, 1973). Given that the ultimate goal of any evaluation is to continue successful programs or improve less successful ones, communication in the forms of public disclosure and evaluator-evaluatee exchange is essential. Indeed, systematic communication should be viewed as a hallmark of sound evaluation.

Climate for an Effective Evaluation

Evaluation conducted in an environment that fosters mutual trust between the evaluator (representing the school) and the teacher holds the greatest potential to benefit both parties. Although teacher evaluation as a process too frequently has tended to generate suspicion, if not outright conflict, trust between evaluator and evaluatee can prevail in an effective personnel evaluation system (Stronge, 1991). Redfern (1980) maintains that a cooperative atmosphere engendered in the evaluation process enhances opportunities for satisfying the concerns of all parties involved.

Satisfaction, a state closely associated with the concept of climate (Owens, 1987), must exist if teacher evaluation systems are to fulfill their potential. A critical variable in creating a climate of satisfaction in the workplace is consideration (i.e., fair and humane treatment among employees and between employer and employee). George (1987) notes that performance management is based on communication and personal relationships and that, as in other relationships, "the qualities of empathy, honesty, and esteem need to be consistent" (p. 23).

Although a supportive climate is logically related to productive evaluation, there are instances when productive evaluation can take place in less than ideal environments. When employees are not measuring up to reasonable performance expectations, and when improvement through remediation has not remedied the problem, then it may be necessary to proceed with meaningful evaluations and, if necessary, to make negative personnel decisions based on the results of those evaluations. It is essential to remember that, ultimately, the best interests of the school's primary clients, the students, must be protected.

Technically Sound Evaluation Systems

A conceptually sound and technically correct teacher evaluation system will not guarantee effective evaluation, but one that is technically flawed and irrational most assuredly will guarantee failure. Evaluation systems that are conceptually and technically sound promote the likelihood that desirable outcomes will be achieved, such as those described in the guiding assumptions of the Personnel Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1988):

- To provide effective service to students and society
- To establish personnel evaluation practices that are constructive and free of unnecessary threatening or demoralizing characteristics
- To facilitate planning for sound professional development experiences

Technically sound evaluation systems enjoy the benefits of the four basic standards espoused in the Personnel Evaluation Standards:

- Propriety (i.e., they are legally and ethically acceptable)
- Utility (i.e., they are useful, informative, timely, and influential)
- Feasibility (i.e., they are efficient, viable in the context of the organization, and relatively easy to use)
- Accuracy (i.e., they are valid and reliable)

Proper use of the Personnel Evaluation Standards can provide assurances of quality control to stakeholders and can support improvement in the overall personnel evaluation process (Stufflebeam & Brethower, 1987; Stufflebeam & Sanders, 1990).

Use of Multiple Data Sources

An important feature of an effective teacher evaluation system is the use of multiple data sources for documenting performance. The most common method for evaluating teachers is based on a clinical supervision model consisting of preconference, observation, and postconference. In fact, as noted in a study conducted by the Educational Research Service (1988), 99.8% of American public school administrators use direct classroom observation as their primary data collection technique. However, primary reliance on formal observation in evaluation presents significant problems (e.g., artificiality, small sample of performance).

The creative use of multiple data sources to measure teacher performance can result in a fuller and more accurate view than is available through more narrowly defined approaches to data collection (Helm, 1994). Although formal classroom observation can provide significant data, it is too limited to be used as the sole source of data for teacher evaluation.[3] Teacher performance can best be judged through more comprehensive and inclusive means.

The use of multiple data sources in the teacher evaluation process offers numerous advantages over single-source data collection (see, for example, Conley, 1987; Glatthorn, 1984; Harris, 1987; McGreal, 1988):

- Production of a more complete and richly textured portrait of the evaluatee's performance
- Collection of data in more naturally occurring situations
- Integration of primary and secondary data sources in the evaluation
- Assurance of greater reliability in the documentation of performance
- Enhancement of objectivity in the documentation of performance
- Documentation of performance that is more closely related to actual work
- A more legally defensible basis for evaluation decisions

Major data sources for teacher evaluation include the following:

- Observation (i.e., formal classroom/work setting observation and ongoing anecdotal observation of performance)
- Client feedback (i.e., client interviews or surveys of students, parents, subordinates, and peers)
- Student performance data (i.e., student achievement)
- Portfolios (i.e., actual materials and reflections on performance logs, case notes, lesson plans)
- Self-evaluation (i.e., self-reflection and analysis of performance)[4]

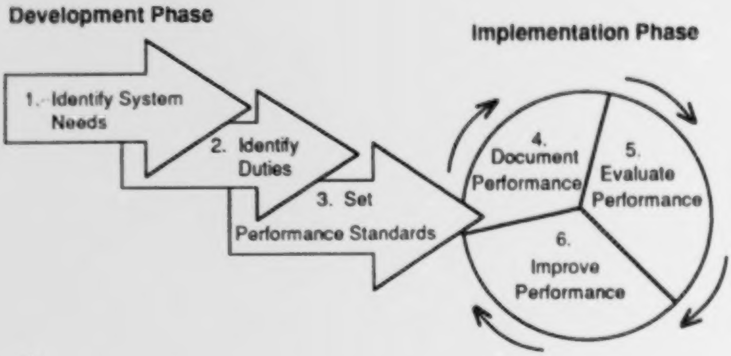

Development Phase
Implementation Phase
1. Identify System Needs
2. Identify Duties
3. Set Performance Standards
4. Document Performance
5. Evaluate Performance
6. Improve Performance

Figure 1.1. Goals and Roles Evaluation Model
SOURCE: Stronge et al. (1995, p. 8). Copyright 1995 James H. Stronge. Reprinted by permission.

The integration of multiple data sources in a teacher assessment and evaluation system offers a much more realistic picture of actual job performance and provides a stronger platform upon which to build realistic improvement plans than would be possible based on only a single source of information such as classroom observation. As multiple data sources are properly employed in performance evaluation, the validity and utility of the process can be dramatically enhanced.

How Can a Teacher Assessment and Evaluation System Be Designed for School Improvement and Personal Growth?

If a teacher assessment and evaluation system is to be effective, it is important that evaluators consider the unique contributions made by each teacher to the accomplishment of the school's mission. The goals and roles evaluation model (Stronge, 1988, 1995; Stronge & Helm, 1990, 1991, 1992)[5] is a practical, research-based model of teacher evaluation that is rooted in the premise of individual-institutional improvement (see Figure 1.1). The model is designed generically for use with a variety of positions, and it may serve well as the basis for an evaluation system not only for teachers but for administrators and support personnel as well.

The goals and roles evaluation model reflects two phases in the evaluation process and involves six distinct steps:

- Development phase
  1. Identify system needs.
  2. Develop roles and responsibilities for the job.
  3. Set standards for job performance.
- Implementation phase
  4. Document job performance.
  5. Evaluate performance.
  6. Improve/maintain professional service.

Each of the six steps is described briefly below.

Development Phase

Step 1: Identify System Needs

Each school has specific needs that relate to its mission and that are met through the collective performance of all personnel (e.g., principal, classroom teachers, resource specialists, counselors). A systematic examination of the needs of the school's constituents will help clarify its mission and purpose. Goals should be developed within the context of the greater community and in consideration of relevant variables, such as financial and personnel resources. School- or districtwide goals often are found in a mission statement, a set of educational goals, a multiyear school plan, or a strategic plan.

Once school goals have been established, attention should turn to the matter of translating those goals into operational terms. One logical way of accomplishing this task is to consider all programs (e.g., math curriculum, guidance counseling services, athletic program) in light of whether they contribute to the accomplishment of the school's goals, and then relate program objectives to position expectations (Stronge & Tucker, 1995). In essence, a domino effect is initiated in this line of planning and evaluation:

→ School goals dictate which programs are relevant to the school's mission,
  → program objectives dictate what positions are needed in order to fulfill those objectives, and finally,
    → individual teacher positions are implemented (and, ultimately, evaluated) in light of the duties and responsibilities dictated by the program objectives.

Determining the needs of the organization is prerequisite to all the remaining steps if the evaluation process is to be relevant to the school's mission and, ultimately, responsive to public demands for accountability (Castetter, 1981; Connellan, 1978; Goodale, 1992, Locke, 1968; Patton, 1986; Phi Delta Kappa National Study Committee on Evaluation, 1971).

Step 2: Develop Teacher Roles and Responsibilities

Accurate and appropriate descriptions of the teacher's roles and responsibilities can be developed only from clear statements of school or district goals and philosophies. Once school goals are determined, then it is only sensible to relate program expectations to position expectations. Typical areas of responsibility might include instructional planning and delivery, assessment, management, and professionalism. These areas of responsibility can serve as a framework for the categorization of more specific responsibilities (or "duties") (Olsen & Bennett, 1980; Redfern, 1980; Scriven, 1988a, 1988b, 1991).[6]

Because job performance must reflect behavior in order to be evaluated, an important addition to the definition of the teacher's role and responsibilities is to identify sample performance indicators. Professional responsibilities are intended to capture the essence of the teacher's job, but it is difficult, if not impossible, to document the fulfillment of professional responsibilities without some measurable indication of their accomplishment. Thus, to give meaning to the teacher's professional responsibilities, it is advisable for the evaluator to select a *sampling* of performance indicators that are both measurable and indicative of the job (Bolton, 1980; Keegan, 1975; Redfern, 1980; Valentine, 1992).

Step 3: Set Standards for Job Performance

Setting standards involves determining a level of acceptable performance. It also may entail determining performance that exceeds acceptable expectations. Because of school needs, available resources, the purpose of a specific position, and a variety of other factors, standards of performance may vary from organization to organization. Although operational definitions for standards may vary from organization to organization, they must be standard and consistently implemented within a school or school district in order to ensure fairness and legal defensibility. This step is important in any goals-oriented personnel evaluation system and should be addressed by both the administrator-evaluator and the teacher (Joint Committee on Standards for Educational Evaluation, 1988; Manatt, 1988; Phi Delta Kappa National Study Committee on Evaluation, 1971; Scriven, 1988b)

Steps 1 through 3 constitute the development phase of the goals and roles evaluation model and are intended to be given careful consideration in advance of the implementation phase.

Implementation Phase

Step 4: Document Job Performance

Documentation is the process of recording sufficient information about the teacher's performance to support ongoing evaluation and to justify any personnel decisions based on the evaluation. The basic question is: How will the teacher demonstrate performance of the roles and responsibilities of the job? Documentation procedures need to rely on multiple data sources, as discussed above (Conley, 1987; Ellett, 1987; Glatthorn, 1984; Harris, 1987; Helm, 1994; McGreal, 1988).

Step 5: Evaluate Performance

Evaluation is the process of comparing an individual teacher's documented job performance with previously established roles and responsibilities and acceptable performance standards. Although this step clearly entails an end-of-cycle summative evaluation, evaluating performance also must include periodic feedback through formative assessment. By providing the teacher with feedback throughout the evaluation cycle, the administrator-evaluator supports the teacher in his or her ongoing efforts to fulfill performance expectations and helps to identify areas of performance that need attention while there is still time to improve. Additionally, opportunity for adequate notice is provided through periodic formative feedback, leading to a fair summative evaluation in which there should be no surprises.

Summative evaluation provides an opportunity for the evaluator to determine individual merit based on performance. Further, the evaluation affords the basis for judging worth, first, by allowing evaluation performance to be viewed in light of the school's goals and, second, by making possible the maintenance of compatibility between individual performance and school goals. In an ongoing, systematic evaluation process, the identification of system needs and the relating of those needs to performance ensure that the evaluation is concerned with both the merit (internal value) and the worth (external value) of performance (Bridges, 1990a, 1990b; Frels & Horton, 1994; Medley, Coker, & Soar, 1984; Scriven, 1973, 1995; Valentine, 1992).

Step 6: Improve/Maintain Professional Service

With an emphasis in the evaluation process on both improvement (i.e., formative) and accountability (i.e., summative) purposes, this final step brings the goals and roles evaluation process full circle. Formative aspects of evaluation, intended to provide recognition for noteworthy performance along with immediate and intermediate feedback for performance improvement and correction where needed, should be ongoing throughout the evaluation process and are implicit in this model. Nonetheless, it is beneficial to include an explicit step aimed at improving and/or maintaining professional service as the culmination of the evaluation cycle and as an entrée into the following cycle.[7]

Improving and maintaining professional service may take the form of a variety of personnel decisions, including assisting the teacher in improving performance, personnel transfers, and, when necessary, termination. Moreover, this step suggests the importance of professional development with a balance between the interests of the teacher and the interests of the school, in a continuous improvement cycle (Little, 1993). After all, the most fundamental purpose of an evaluation is to *improve* both the individual's and the institution's performance (Hunter, 1988; Iwanicki, 1990; McGreal, 1988; Stufflebeam, 1983).

Although I have described six distinct steps in the goals and roles evaluation model, it is important to recognize that the first three steps require intensive attention only during the development phase. For example, once the organizational goals are established (Step 1), it is not necessary to reassess those goals every year; rather, most schools implement a mid- to long-range systematic plan that is reviewed and updated periodically. Additionally, once roles and responsibilities are determined for a given position (Step 2) and performance standards are established (Step 3), there typically is need only for review and minor modifications during the implementation phase. Thus, once the development phase has been completed, most of the energy and effort are invested in the implementation phase (the last three steps of the evaluation): documenting performance (Step 4), evaluating performance (Step 5), and improving and maintaining professional service based on that performance (Step 6).

Summary

Effective teacher evaluation is essential for effective schools. Among the advantages of a high-quality teacher evaluation system are the following:

- Administrators and teachers are jointly involved in the design of the evaluation system.
- The entire professional staff of the school is included in a comprehensive evaluation system.
- School goals and individual responsibilities are rationally linked.
- The objectives of evaluation for the teacher are clearly established.
- A basis for objective evaluation is provided.
- Systemwide resources are more efficiently channeled into a unified evaluation process.

- Evaluator training is made more manageable and meaningful through a unified approach to evaluation for all employees.
- Performance feedback based on observable, job-related results is provided.
- All professional employees, including teachers, are provided appropriate systematic opportunities for improvement.
- A school is more accountable to its public when it implements a comprehensive evaluation system that includes in a meaningful way *all* of its professional employees.
- The treatment of teachers and others under such a system is legally defensible.

In the final analysis, teacher evaluation in its myriad forms is nothing more than a process for determining how an individual, program, or school is performing in relation to a given set of circumstances. When evaluation is viewed as more than this process (i.e., evaluation as an end in itself), it gets in the way of progress and thus becomes irrelevant. When evaluation is treated as less than it deserves (i.e., superficially, with little or no resource allocation, invalid evaluation systems, and flawed implementation designs), the school, its employees, and the public at large are deprived of opportunities for improvement and the benefits that accountability can afford. All of us, whatever our relationship to the educational enterprise, deserve high-quality evaluation. A teacher assessment and evaluation system that is built squarely upon individual and institutional improvement holds the promise of filling this need and better serving our students and our communities.

Notes

1. The term *assessment* is typically used to connote the formative aspects of performance review, whereas *evaluation* is sometimes used to describe the summative aspects. Although I use both terms in this chapter, *evaluation* is considered to be the more inclusive term; thus, it is used here to include both formative and summative aspects of performance review.
2. See Scriven (1967) for a discussion of the concepts of formative and summative purposes in evaluation.
3. For additional discussion of the limits of classroom observation for teacher evaluation, see Chapter 6 of this volume.
4. Separate chapters are provided in this text on each of these data sources.

5. Initial work on the model took place in the context of evaluation for educational specialists (e.g., school counselor, school psychologist) described as professional support personnel. The model is intended to provide an evaluation paradigm that can be adapted to any individual and institution.
6. Scriven (1991) has developed an extensive list of the duties of the teacher (see also Chapter 2, this volume).
7. The length of an evaluation cycle should be determined based on the context of the setting. In some schools and for some teachers (e.g., those on probationary status), a 1-year cycle may be most beneficial; in other settings, a multiyear cycle may work best.

References

Bolton, D. L. (1980). *Evaluating administrative personnel in school systems*. New York: Teachers College Press.

Bridges, E. M. (1990a). Evaluation for tenure and dismissal. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 147-157). Newbury Park, CA: Sage.

Bridges, E. M. (1990b). *Managing the incompetent teacher*. Eugene, OR: ERIC Clearinghouse on Educational Management.

Castetter, W. B. (1981). *The personnel function in educational administration* (3rd ed.). New York: Macmillan.

Clark, D. L., & Astuto, T. A. (1994). Redirecting reform: Challenges to popular assumptions about teachers and students. *Phi Delta Kappan, 75*, 513-520.

Conley, D. T. (1987). Critical attributes of effective evaluation systems. *Educational Leadership, 44*(7), 60-64.

Connellan, T. K. (1978). *How to improve human performance: Behaviorism in business and industry*. New York: Harper & Row.

Cummings, L. L., & Schwab, D. P. (1973). *Performance in organizations: Determinate and appraisal*. Glenview, IL: Scott Foresman.

Duke, D. L. (1990). Developing teacher evaluation systems that promote professional growth. *Journal of Personnel Evaluation in Education, 4*, 131-144.

Duke, D. L. (1993). Removing barriers to professional growth. *Phi Delta Kappan, 74*, 702-712.

Educational Research Service. (1988). *Teacher evaluation: Practices and procedures*. Arlington, VA: Author.

Ellett, C. D. (1987). Emerging teacher performance assessment practices: Implications for the instructional supervision role of school principals. In W. Greenfield (Ed.), *Instructional leadership: Concepts and controversies* (pp. 302-327). Boston: Allyn & Bacon.

Frels, K., & Horton, J. L. (1994). *A documentation system for teacher improvement and termination*. Topeka, KS: National Organization on Legal Problems in Education.

Fullan, M. G. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Fullan, M. G. (1996). Turning systemic thinking on its head. *Phi Delta Kappan, 77,* 420-423.

George, P. S. (1987). Performance management in education. *Educational Leadership, 44*(7), 32-39.

Glatthorn, A. (1984). *Differentiated supervision*. Alexandria, VA: Association for Supervision and Curriculum Development.

Goodale, J. G. (1992). Improving performance appraisal. *Business Quarterly, 51*(2), 65-70.

Harris, B. M. (1987). Resolving old dilemmas in diagnostic evaluation. *Educational Leadership, 44*(7), 46-49.

Helm, V. M. (1994, April). *The professional support personnel evaluation model: The use of multiple data sources in personnel evaluation*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Hunter, M. (1988). Create rather than await your fate in teacher evaluation. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 32-54). Alexandria, VA: Association for Supervision and Curriculum Development.

Iwanicki, E. F. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook for teacher evaluation: Assessing elementary and secondary school teachers* (pp. 158-171). Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to assess systems of evaluating educators*. Newbury Park, CA: Sage.

Keegan, J. J. (1975). Performance based staff evaluation: A reality we must face. *Educational Technology, 15,* 35-38.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15,* 129-151.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3,* 157-189.

Manatt, R. P. (1988). Teacher performance evaluation: A total systems approach. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 79-108). Alexandria, VA: Association for Supervision and Curriculum Development.

March, J. G., & Simon, H. A. (1967). *Organizations.* New York: John Wiley.

March, J. G., & Simon, H. A. (1993). *Organizations* (2nd ed.). Cambridge, MA: Blackwell Business.

McGaghie, W. C. (1991). Professional competence evaluation. *Educational Researcher, 20,* 3-9.

McGreal, T. L. (1988). Evaluation for enhancing instruction: Linking teacher evaluation and staff development. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 1-29). Alexandria, VA: Association for Supervision and Curriculum Development.

McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher evaluation: Improvement, accountability, and effective learning.* New York: Teachers College Press.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance.* New York: Longman.

Olsen, L. D., & Bennett, A. C. (1980). Performance appraisal: Management technique as social process? In D. S. Beach (Ed.), *Managing people at work: Readings in personnel* (3rd ed.). New York: Macmillan.

Owens, R. G. (1987). *Organizational behavior in education* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Beverly Hills, CA: Sage.

Phi Delta Kappa National Study Committee on Evaluation. (1971). *Educational evaluation and decision making.* Itasca, IL: F. E. Peacock.

Pogrow, S. (1996). Reforming the wannabe reformers: Why education reforms almost always end up making things worse. *Phi Delta Kappan, 77,* 656-663.

Redfern, G. B. (1980). *Evaluating teachers and administrators: A performance objectives approach.* Boulder, CO: Westview.

Saphier, J. (n.d.). *How to make supervision and evaluation really work: Supervision and evaluation in the context of strengthening school culture.* Carlisle, MA: Research for Better Teaching.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

Scriven, M. (1972). The methodology of evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education.* Boston: Allyn & Bacon.

Scriven, M. (1973). The methodology of evaluation. In B. R. Worthen & J. R. Sanders (Eds.), *Educational evaluation: Theory and practice.* Belmont, CA: Wadsworth.

Scriven, M. (1988a). Duties-based teacher evaluation. *Journal of Personnel Evaluation in Education, 1,* 319-334.

Scriven, M. (1988b). Evaluating teachers as professionals: The duties-based approach. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 110-142). Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1991). *Duties of the teacher* (TEMP A memo). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation, 21,* 111-129.

Stronge, J. H. (1988). *Evaluation of ancillary school personnel training module—Illinois Administrators Academy.* Springfield: Illinois State Board of Education.

Stronge, J. H. (1991). The dynamics of effective performance evaluation systems in education: Conceptual, human relations, and technical domains. *Journal of Personnel Evaluation in Education, 4,* 405-411.

Stronge, J. H. (1993). Evaluating teachers and support personnel. In B. S. Billingsley (Ed.), *Program leadership for serving students with disabilities* (pp. 445-464). Richmond: Virginia Department of Education.

Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation, 21,* 131-151.

Stronge, J. H., & Helm, V. M. (1990). Evaluating educational support personnel: A conceptual and legal framework. *Journal of Personnel Evaluation in Education, 4,* 145-156.

Stronge, J. H., & Helm, V. M. (1991). *Evaluating professional support personnel in education.* Newbury Park, CA: Sage.

Stronge, J. H., & Helm, V. M. (1992). A performance evaluation system for professional support personnel. *Educational Evaluation and Policy Analysis, 14,* 175-180.

Stronge, J. H., Helm, V. M., & Tucker, P. D. (1995). *Evaluation handbook for professional support personnel.* Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Stronge, J. H., & Tucker, P. D. (1995). Performance evaluation of professional support personnel: A survey of the states. *Journal of Personnel Evaluation in Education, 9,* 123-138.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services in evaluation* (pp. 117-141). Boston: Kluwer-Nijhoff.

Stufflebeam, D. L., & Brethower, D. M. (1987). Improving personnel evaluations through professional standards. *Journal of Personnel Evaluation in Education, 1,* 125-155.

Stufflebeam, D. L., & Sanders, J. R. (1990). Using the Personnel Evaluation Standards to improve teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 416-428). Newbury Park, CA: Sage.

Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research, 35,* 492-501.

Valentine, J. W. (1992). *Principles and practices for effective teacher evaluation.* Boston: Allyn & Bacon.

PART I

Designing a Teacher Evaluation System

2

Building the Foundation: Teacher Roles and Responsibilities

PATRICIA H. WHEELER
MICHAEL SCRIVEN

When building a teacher evaluation system, the developers are usually aware of the purpose—for example, for accountability or improvement, or a mixture of these. However, once the purpose is specified, the next steps are often taken too quickly. Although it is crucial to write criteria, set standards, design assessments, and prepare procedures, the next step should be the selection of an appropriate foundation upon which to build a valid system. The foundation provides "the rationale for the attributes and domains covered by a teacher evaluation system" (Wheeler, 1995, p. 3). Possible foundations include research on teaching, government policies, professional expertise, theories of teaching and learning, and teacher duties and professional responsibilities.

Once an appropriate type of foundation is identified and tentatively specified, the foundation should undergo a careful review—a descriptive or prescriptive content analysis with substantial input by teachers and administrators. Then the next steps of writing criteria and indicators, determining standards for various levels of performance, selecting and/or designing assessment instruments, and preparing evaluation procedures can begin. These further steps are just as important as identifying the foundation. The use of a suitable foundation is a necessary, but not a sufficient, condition for a valid and legally defensible system. It also helps to ensure that the criteria reflect the school's and district's missions and policies.

The failure to use an appropriate and a solid foundation can have serious consequences for the teacher evaluation system and, hence, for the educational services provided by the school. For example, the criteria may be incomplete or may include indefensible requirements, or the standards may be too low or too high, in which case good teachers will be overlooked and poor ones retained and promoted.

Purposes of Teacher Evaluation Systems

The purpose of a particular teacher evaluation system is one key to selecting the appropriate foundation upon which to build that system. There are a multitude of purposes for evaluating teachers, some of which are discussed in Chapter 1 of this volume. Possible purposes include the following:

- Hiring
- Salary decisions
- Assigning
- Reduction in force
- Performance evaluation
- Retirement exemption
- Pretenure retention/termination
- Licensing/credentialing
- Tenure
- Awards/recognition
- Posttenure retention/termination
- Self-assessment
- Promotion/career ladder
- Mentoring appointment

Ten purposes for evaluating teachers are identified in the Personnel Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (1988): entry to training, certification/licensing, defining a role, selection, performance reviews, counseling for staff development, merit awards, tenure decisions, promotion decisions, and termination.

Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) list four basic purposes of teacher evaluation, the first two focusing on improvement and the last two focusing on accountability: individual staff development, school improvement, individual personnel decisions, and school status decisions. Wheeler (1991) identifies several uses of teacher evaluation for personnel decisions: hiring, job assignments, formal evaluations, retention and termination, tenure, salary and other compensation, career ladder and promotion, reduction in force, and exemption from age-based retirement. The purposes of the teacher evaluation system must be clearly stated and the users and uses delineated before the determination is made regarding which foundations to use.

What Are Some Possible Foundations?

Just as there are multiple purposes for a teacher evaluation system, there are several types of foundations upon which a system can be built (Wheeler, 1991, 1992a, 1994b, 1995). The foundations covered in this chapter are of six types:

- What teachers are doing or say they are doing as part of their job
- What administrators or others would like teachers to be doing
- What research says effective teachers do
- What theories of teaching and learning say teachers should do
- What the outcomes of teaching are
- What the roles and responsibilities of teachers are

Examples of each of these types of foundations are provided below, and some of the problems associated with each are delineated.

What Teachers Are Doing or Say They Are Doing as Part of Their Job

One type of foundation relies on input from teachers about what they are doing as part of their job. Examples of these are job analysis studies, teacher surveys, teacher consensus about the nature of their job, logs or self-reports of individual teachers about what they do, and the teaching norms and traditions of the school.

Job analysis is a type of research that focuses on the job components of what a teacher is doing. It identifies the tasks and activities as well as the knowledge, skills, and abilities needed to perform those tasks. It can be based on extensive and comprehensive observations of teachers at work or on surveys of teachers about their jobs. An example of a job analysis that looked at the coverage of an assessment for potential teachers was that conducted for the National Teacher Examination (NTE) Core Battery by Rosenfeld, Thornton, and Skurnik (1986). These researchers identified six major professional functions of teachers:

1. Managing and influencing student behavior
2. Clerical, administrative, and other professional functions
3. Assessing, grading, and recording student learning progress and evaluating instructional effectiveness
4. Planning the lessons, selecting the materials, and previewing the instructional program
5. Implementing the planned instructional program, using a variety of techniques
6. Identifying students with individual or similar instructional needs and teaching them accordingly

Based on their own experience, teachers can be asked to delineate their own job functions and professional responsibilities, and to identify those areas that affect how a person performs the job of teaching. This can be done in a focus group, to build consensus among the teachers on the nature of their jobs; through self-reports by individual teachers, which later need to be compiled; or by ascertaining the teaching norms and traditions of the school through a close look over time at how the school functions.

What Administrators or Others Would Like Teachers to Be Doing

The second kind of foundation involves what administrators and others say concerning how teachers should do their job. This is sometimes referred to as style-based evaluation. Examples of evaluation systems based on such foundations are programs requiring the use of certain teaching practices or strategies, the imposition of specific styles on teachers by other educators or policy makers, and the preferences of those doing the evaluating as well as other individuals or groups.

A teacher evaluation system using this type of foundation focuses on a certain style of or approach to teaching. The criteria reflect what a teacher using that style of teaching ought to be doing. Examples are cooperative learning, whole language instruction, lecturing, and team teaching. The assumptions are that a particular teaching style is best for the students in the school and that only teachers who use that style well are good teachers. Ornstein (1991) points out that such assumptions are questionable: "Teacher style is a matter of choice and comfort, and what works for one teacher with one set of students may not work for another" (p. 72).

This foundation for evaluation may be appropriate when a school is trying to implement a new instructional program or strategy and wants to see how well each teacher is doing with that style. However, it is not an appropriate foundation for most teacher evaluation systems and should not be used for personnel decisions. As Scriven (1994b) reminds us, "The imposition of style requirements, even if they were based on good evidence overall, imposes a handicap on many individual teachers and hence on the children" (p. 87).

The professional judgment of a group of experts is commonly used as the foundation for a teacher evaluation system. As a group or individually, these experts identify the criteria and indicators that reflect what they regard as good teaching. Such experts typically include experienced teachers, curriculum specialists, administrators, teacher trainers, other college faculty, and researchers. Additionally, some professional organizations offer documents that can be helpful in the development of teacher evaluation systems (e.g., American Association of School Personnel Administrators, 1995; National Council of Teachers of Mathematics, 1991).

BEST COPY AVAILABLE

One can also ask the clients of teachers—the students, parents, community members, local business leaders, and others—what makes a teacher a good teacher. They will tend to base their responses on the array of teacher behaviors to which they have been exposed throughout their own schooling, their own educational needs, and comments about teachers they have heard from others. Although this is not likely to be a strong foundation by itself, it can help reinforce other foundations and, at times, make one aware of areas that may have been overlooked. A teacher evaluation system could be built, in part, based on the judgments of these various client groups and stakeholders about what the results of a teacher's work should be. There is apt to be more acceptance of the system if this is one of the foundations used; at least, these stakeholder groups will have an opportunity to have input on what criteria the system covers.

What Research Says Effective Teachers Do

Another foundation for a teacher evaluation system involves what teachers are doing. An evaluation system can be built based on the results of studies of effective and ineffective teachers that identify the practices and behaviors associated with these teachers (effective teaching research, or process-product studies).

Research on effective teaching has identified a number of teaching practices that are associated with desired results (e.g., Brophy & Good, 1986). A comprehensive review of research on effective teaching has frequently been used as a foundation for teacher evaluation systems. However, such a foundation should be used with extreme caution. As McDonald and Elias (1976) found in the Beginning Teacher Evaluation Study, what worked well in one subject or at one grade level or in one teaching context did not necessarily work well in another. The generalizability of findings from research on effective teaching is questionable; research alone should not be a basis for building a teacher evaluation system. As Scriven (1990) points out, "Many researchers who have encouraged the use of this research by schools and colleges—and states and districts—as a basis for a teacher evaluation system have seriously misled their readers and clients" (p. 19).

Shuell (1993), based on his premise that teaching and learning must be addressed simultaneously, identifies three considerations that should be taken into account in the use of such research. First, few studies look at "the relationship between specific teaching practices and the consequent learning processes elicited in students" (p. 302). Second, often such studies look at learning and cognition in isolation from other concerns, such as metacognitive, affective, motivational, and developmental processes. And finally, "research on teaching and learning should be concerned with teacher evaluation content and context of learning as well as the process of learning" (p. 302).

Research findings are useful in preservice teacher training, inservice staff development, and the review of a teacher evaluation system to ensure that no key areas or criteria and performance indicators have been overlooked. However, effective teaching research and process-product studies should not be a primary foundation for building a teacher evaluation system.

What Theories of Teaching and Learning Say Teachers Should Do

Scriven (1991) has defined theories as "general accounts of a field of phenomena, generating at least explanations and sometimes also predictions and generalizations" (p. 360). The emphasis in theories is on explanations. These may be based on styles or preferences of key leaders in education or on the findings of research. However, Scriven points out that explanatory power is not the only indicator of a sound theory (and thus solid foundations). Other aspects to consider include "predictive power and accuracy, economy of assumptions, fertility of implications, cross-field support (e.g., by analogy), and simplicity of claims" (p. 360).

Theories that are likely to inform teacher evaluation systems are described as middle-range theories grounded in practice. They provide theoretical propositions that shape inquiry and generate further research, but are more limited in scope and longevity than the type of theory described by Scriven (1991). Middle-range theories that could be considered as foundations for teacher evaluation systems include theories of teaching, cognitive theories of teachers' thought processes, and theories of the cognitive development of teachers.

Examples of theories of teaching are provided by Joyce and Weil (1986), who describe four models of teaching, each justified by a theory and linked to outcomes of teaching—social interaction models, information processing models, personal models, and behavior modification models. In their theory of teaching, Medley, Coker, and Soar (1984) propose three levels of teaching—environmental maintenance, implementation of instruction, and individualization—all of which occur simultaneously in a classroom.

Theories in the cognitive psychology of teaching look at teachers' thought processes. They cover such areas as planning, decision making, judgment, implicit theories, expectations, and attributions (Clark & Peterson, 1986). Cognitive psychologists have developed theories of the cognitive development of teachers. Costa, Garmston, and Lambert (1988), for example, offer a set of benchmarks that teachers should attain with training and experience. The benchmarks are defined by three stages of teacher development (entry, tenure, master) and four cognitive development areas (learning, doing, valuing, thinking). Such theories are not, by themselves, sufficient as a valid foundation for a teacher evaluation system, although they can supplement other foundations.

What the Outcomes of Teaching Are

As is often done for summative evaluation of educational programs, teachers are sometimes evaluated based on the extent of their attainment of a set of goals and objectives, and on the outcomes of their teaching. Examples include management by objective, professional development goals, student outcomes, teacher productivity, and client and stakeholder satisfaction.

With management by objective, the criteria for the evaluation are based on the contributions of each teacher to the overall objectives and common goals delineated by the school. The goals and objectives are set by a local policy group (e.g., the school board, the school site council) or by the school or district administration, although some teachers may be involved in the setting of goals and objectives.

As part of their own professional growth, teachers may be asked to set developmental goals for themselves. These goals are usually written with the assistance of an administrator or a mentor teacher, and are agreed on by both parties. Each teacher is evaluated based on his or her own goals. This foundation is more frequently used for formative than for summative evaluation.

An array of student outcomes can be used as a foundation to evaluate teachers. These include direct measures of student learning and development (behavioral, physical, social). The rationale behind a student outcomes foundation is that "student learning is both the goal and the product of teaching" (Robinson, 1984, p. 15). Standardized tests are the most frequently used student outcome measures for evaluating teachers, although such use of these tests can have serious limitations (e.g., Berk, 1988; Haertel, 1986; Millman, 1981; Wheeler, 1995-1996).

Sometimes other student outcome measures are used, such as measures of academic, physical, social, and personal development. Examples include grades, promotion and graduation rates, performance on physical tasks (e.g., pull-ups, fine motor coordination tasks), referrals for disciplinary action, attendance, and service hours to the school. Although such indicators may seem objective, they can be difficult to link to the performance of an individual teacher, and they reflect only part of a teacher's array of responsibilities and of the content and skills taught to students by that teacher.

Teacher productivity reflects the accomplishments of a teacher during the period covered by the evaluation. The measures used most often in elementary and secondary schools for teacher evaluation are those listed above for student outcomes. Examples of other productivity variables include number of students taught, volume of content covered or number of activities completed in a given period of time, number of articles and books published, amount of resources used, and cost per student. Productivity approaches to teacher evaluation tend to quantify teaching.

At the postsecondary level, student ratings of teacher performance are often used to evaluate faculty (Braskamp, Brandenburg, & Ory, 1984; Centra, 1980). This approach is based on a client satisfaction foundation. At the elementary and secondary levels, not only are students a potential source of information about what teachers should do to satisfy their needs, but so are other stakeholders—parents, community agencies, local employers. A teacher evaluation system could be based, in part, on input from these various client groups. It must be kept in mind that outcomes-based teacher evaluation systems "can drive instruction and teaching behavior rather than promote diverse instructional approaches and curricular content for different teachers and students" (Wheeler, 1994c, p. 3).

What the Roles and Responsibilities of Teachers Are

Rather than building a teacher evaluation system on a foundation of what teachers are doing or what someone thinks they should be doing, it is more appropriate to consider what teachers are hired to do—that is, their professional roles and responsibilities in given positions. This includes their professional responsibility to fulfill entry-level requirements for the job and to keep current on fulfillment of requirements.

The teacher's roles and responsibilities, including entry-level requirements, should be delineated in a job description or similar document at the time of employment by the school (McConney et al., 1996). Professional roles and responsibilities can include such areas as being knowledgeable in the subject matter, earning and maintaining current teaching credentials, reviewing and selecting curriculum materials, designing instruction and planning lessons, monitoring and assessing student learning, communicating with parents, maintaining records of student learning, fulfilling applicable laws and government regulations, and participating in professional service and staff development activities.

Evaluation systems for in-service teachers should have built-in procedures for monitoring ongoing fulfillment of applicable licensing and credentialing requirements. Teachers have a professional responsibility to maintain minimal levels for such requirements over time. For example, although when a teacher was licensed there may have been evidence that he or she had a minimally acceptable level of subject-matter knowledge, changes in the field or years of not teaching certain subject areas might indicate the necessity for a teacher evaluation system to monitor subject-matter knowledge with regard to licensure requirements as part of evaluation and job assignment. Or state licensing requirements may have been updated, and changes may apply to all teachers, including those currently teaching.

In addition to licensure and credentialing requirements, there may be other governmental requirements that affect public employees, including teachers. Usually such regulations are designed to protect the public. For example, they might be concerned with criminal records, federal or state safety codes (e.g., storage of chemicals in the science lab, proper use of safety equipment in the wood shop), or the need for special training (e.g., first aid/CPR for physical education teachers and coaches). Such licensing and governmental requirements should be addressed by a teacher evaluation system, because they are directly related to the roles and responsibilities of teachers.

Scriven (1988b) refers to this type of teacher evaluation system as duties-based teacher evaluation, or DBTE. Figure 2.1 contains an outline of the duties of the teacher, or DOTT (Scriven, 1988a, 1988b, 1994a), which can provide a good starting point for building an evaluation system on a foundation of roles and responsibilities.

How Does One Choose a Foundation?

When choosing a foundation upon which to build a teacher evaluation system, one must consider some of the problems associated with the major types of foundations. Several such problems (Wheeler, 1995) are shown in Table 2.1. The choice of the wrong foundation can have serious consequences. First, the attributes of teaching and the areas covered by the evaluation system may be incomplete and/or poorly defined. Second, the criteria and standards will lack validity and will not provide a justifiable basis for career actions and personnel decisions regarding teachers.

The problems with the sixth type of foundation, roles and responsibilities, are weaknesses that should be overcome by a school or district before it develops a teacher evaluation system, whereas many of the problems listed for the other five types of foundations cannot be easily overcome or are, in and of themselves, inherent weaknesses that can lead to an invalid system. Despite the potential and real problems with these types of foundations, all foundations can be useful in some aspect of teacher evaluation by providing enlightening perspectives for some of the evaluation tasks or by helping to ensure that key areas are not overlooked.

A system designed to evaluate a teacher's performance on the job would be invalid if it did not address roles and responsibilities. That type of foundation is the keystone for a system used to make personnel decisions, such as retention, dismissal, promotion, transfer, or compensation. The other five types of foundations can be useful for such purposes as professional development, teacher self-evaluation, teacher preparation, licensure and credentialing,

1. **Knowledge of subject matter**
    - A. In the fields of special competence
    - B. In across-the-curriculum subjects
2. **Instructional competence**
    - A. Communication skills
    - B. Management skills
        - i. Management of process
        - ii. Management of progress
        - iii. Management of emergencies
    - C. Course construction and improvement skills
        - i. Course planning
        - ii. Selection and creation of materials
        - iii. Use of available resources (a. local; b. media; c. specialists)
        - iv. Evaluation of course, teaching, materials, and curriculum
3. **Assessment competence**
    - A. Knowledge about student assessment
    - B. Test construction/administration skills
    - C. Grading/ranking/scoring practices
        - i. Process
        - ii. Output
    - D. Recording and reporting student achievement
        - i. Knowledge about reporting achievement
        - ii. Reporting process (to: a. students; b. administrators; c. parents; d. others)
4. **Professionalism**
    - A. Professional ethics
    - B. Professional attitude
    - C. Professional development
    - D. Service to the profession
        - i. Knowledge about the profession
5. **Other duties to the school and community**
    - A. Committee work
    - B. Community work
    - C. Special school assignments (e.g., lunchroom, coaching)

Figure 2.1. Outline of the Duties of the Teacher

SOURCE: Michael Scriven, P.O. Box 69, Point Reyes, CA 94956; e-mail scriven@aol.com; fax (415) 663-1913. Comments of all kinds are extremely welcome.

NOTE: An expanded version of this outline is found in Scriven (1994a), where it is prefaced by a lengthy discussion of concepts, use, and so on.

professional certification, and special awards or recognition. Again, however, it is critical to keep in mind what the professional roles and responsibilities of the teacher are and to ensure that the system addresses them, even if it relies heavily on one or more other types of foundations.

How Are Teacher Roles and Responsibilities Determined?

If the primary purpose of the teacher evaluation system is to determine how well each teacher is performing his or her job, then the most solid and valid foundation to use is the sixth: what teachers should be doing—that is, their roles and responsibilities. Ultimately, we must be concerned with the question of whether teachers are doing what they have been hired to do, that is, fulfilling their assigned roles and responsibilities.

The development of a list of roles and responsibilities can be a challenging and time-consuming process. However, it is essential to the development of a sound teacher evaluation system with valid criteria. One place to start is with the job description, which should cover position title, qualifications, general job responsibilities, specific job responsibilities, physical and mental demands, and working conditions. An example of a partial job description for a middle school science teacher is shown in Figure 2.2.

Many schools and districts lack job descriptions for teachers, or, if they have them, they tend to be minimal, generic, incomplete, and out-of-date. Reasonable job descriptions can be developed using some of the foundations discussed earlier in this chapter as well as input from teachers. In addition to improving the comprehensiveness and validity of the teacher evaluation system, the development of thorough job descriptions delineating teachers' roles and responsibilities can result in many other benefits (McConney et al., 1996). Job descriptions help communicate to individual teachers the requirements of their job assignments and those responsibilities for which they will be held accountable. They contain information for preparing job announcements, developing job descriptions for new positions, and informing job applicants about the expectations of

(text continues on p. 43)

TABLE 2.1 Problems With Major Types of Foundations

What teachers are doing or say they are doing as part of their job:

Assumes that what teachers say they are doing is an accurate and comprehensive portrayal of their jobs, but such self-reporting is often incomplete or may overlook critical responsibilities that apply in intermittent years.

Needs teachers who can clearly articulate and describe what they do.

Requires careful interpretation of what teachers say and caution in not over- or underemphasizing various aspects of the job.

Assumes that what teachers are doing is best for students, parents, schools, and communities, although this is not always true.

Assumes that the long-standing norms and traditions of the school reflect the needs of current students and effective use of available resources.

Leads to a tendency for new teachers to rely heavily on what they learned during their teacher training programs rather than what is best for the students assigned to them and the context in which they are now working.

What administrators or others would like teachers to be doing:

Assumes that what works well for one teacher in a given context with certain students works well for other teachers, regardless of teaching context, which is unreasonable.

Evaluates teachers for their style rather than fulfillment of their responsibilities. Teachers should not be penalized for using a particular style unless it violates law or another style is mandated by law or policy, in which case it becomes a responsibility.

Sprouts from the personal preferences of one or more individuals and may not be in the best interest of the teachers or the students, parents, and community.

Can discourage the use of diverse instructional approaches and curricular content that may be more appropriate for certain teachers and students.

Takes away the responsibility of each teacher as a professional educator to determine how best to meet the needs of students in order to promote student learning and improve student behavior effectively.

TABLE 2.1 Continued

Is clearly invalid for evaluating many teachers, even if the approach works well for some teachers.

What the research says effective teachers do:

Assumes that the same variables that correlate positively with what is regarded as good teaching or with student outcomes are themselves sound bases for evaluating teachers.

May overlook some key variables, because the teaching variables examined usually are determined by the researcher or the sponsor of the research.

Is based on research results that imply guilt by association, using designated teacher characteristics and generalizing those qualities associated with certain types of teachers—such as gender, height, ethnicity, college attended, academic major, and primary language—to all teachers.

Encourages the tendency to generalize research findings across teachers and different teaching contexts when it is inappropriate or unjustifiable to do so.

Makes it easy to overlook the fact that the measures for assessing students and teachers yield only estimates of their levels of performance and that often steps are not taken to minimize error of measurement or to consider error factors in the interpretation of results.

Promotes the tendency to infer causality from correlational research when there is no basis for assuming direct or even indirect causality.

When used for career decision and personnel actions, can result in adverse impacts and may lead to decisions and actions that are legally indefensible and/or technically unsound.

What theories of teaching and learning say teachers should do:

As with styles, is often based on the preferences of key leaders in the field rather than on what works well for a given teacher and his or her students.

Is limited to the aspect of teaching and learning under consideration.

(continued)

TABLE 2.1 Continued

Is based on a certain orientation or way of thinking about teaching and learning, such as constructivism or cognitive psychology, that may be inappropriate for the teacher and school and that limits the ways of looking at teaching and learning.

Can be easily overlooked as only being attempts to explain phenomena and can tend to give theories more emphasis than is warranted.

What the outcomes of teaching are:

Limits the primary focus of the teacher evaluation system to the designated outcome variables.

Becomes more difficult to give meaningful and constructive feedback to teachers on their teaching behaviors because evaluation results shed little light on the teaching processes that led to those outcomes.

Is unable to control for all the confounding factors when student outcome measures are used, such as test scores, absentee rates, and misconduct referrals. Examples of such confounders include missing test data, lack of appropriate tests for disabled or non-English-proficient students, irregular test administration and scoring practices, a schoolwide flu epidemic, and loss of key records in a fire.

Is impossible to isolate the impact that a given teacher has had on the learning and behavior outcomes of individual students, due to the many context features beyond the control of the teacher and the school that affect student outcomes.

Encourages teachers to focus their energies on those outcomes that will be assessed rather than the needs of the students and the teachers' responsibilities. Especially in high-stakes situations, use of this type of foundation can lead to such unintended consequences as teaching to the test, various forms of cheating in administration and scoring of tests, and not implementing the student conduct policy if misconduct referrals are used for evaluation, instead keeping disruptive students in class.

May be inappropriate or irrelevant for teachers who are confronted with challenging situations, difficult teaching contexts, or students with extensive behavior and learning problems.

TABLE 2.1 Continued

Leads to basing evaluation criteria on outcomes that do not clarify the role of the teacher and the teacher's responsibilities.

Assumes that attaining the designated outcomes is the primary function of the teacher.

What the roles and responsibilities of teachers are:

Tends to focus on general responsibilities of all teachers and might overlook additional responsibilities or specific ones applicable to a few teachers.

May lack delineated teacher responsibilities—either general ones for all teachers or specific ones for each teaching position—in job descriptions or elsewhere.

Requires fully informing teachers about state, district, and school procedures, policies, and curriculum frameworks, as they will be evaluated for their implementation of these.

Sometimes encourages too much reliance on external standards and guidelines that may be narrowly focused, may reflect the interest of a certain association or political agenda, or may be irrelevant to the local teaching context, rather than on developing local standards and guidelines.

Requires delineation of teacher responsibilities as well as familiarity with pertinent laws and regulations, and with state, district, and program requirements.

---

the position. They offer individual teachers ideas for use in self-evaluation, and can be helpful to supervisors and mentors who must conduct informal monitoring of individual teachers' performance for purposes of formative evaluation and professional development. They give educators and other stakeholders a better grasp on and understanding of the job of teaching and its demands and challenges. They can be one vehicle for informing preservice teacher education programs, in-service training programs, and licensing boards about the nature of teachers' jobs and for helping them identify areas or requirements that they should address. When a personnel decision is appealed or a case is presented before a court of law or an arbitrator,

(text continues on p. 46)

1. *Position Title:* Middle school science teacher.

2. *Qualifications:* State science teacher credential, either life or physical/earth science. A minimum of four college-level courses in both life sciences and physical/earth sciences and at least four upper-division courses in either life sciences or physical sciences. Fluency in standard English; proficiency in a second language desirable, especially Spanish.

3. *Generic Job Responsibilities:*

   The teacher shall participate as a faculty member in the activities of the school in ways that contribute to the attainment of the purposes and goals of the school district.

   The teacher shall design and implement instructional plans for the student group or groups assigned to him or her.

   The teacher shall instruct students in a manner that will enable them to perform their best on the district testing program.

   The teacher shall accept responsibility for continuing enhancement of his or her professional knowledge, skills, and performance.

   The teacher shall make and keep proper records of student work and the progress of students in his or her charge.

   The teacher shall serve as a school-community liaison person in contacts with parents of students and other members of the community.

   The teacher shall be aware of and observe legal mandates, Board policy, administrative procedures, and contractual agreements as they pertain to his or her position.

   The teacher shall perform all other duties of the position as assigned by the principal and his or her designee within the limits agreed upon by contract and the State Education Code.

4. *Specific Job Responsibilities:*

   The teacher shall inventory and maintain science lab equipment, materials, and live specimens.

   The teacher shall store and use chemicals in accordance with state safety codes and Occupational Safety and Health Administration (OSHA) regulations.

Figure 2.2. Sample of a Partial Job Description
SOURCE: McConney et al. (1996).

The teacher shall implement the state's science curriculum framework for grades 6-8.

The teacher shall plan and coordinate the school's annual science fair.

The teacher shall advise the Science Club.

The teacher shall serve on the district's science curriculum committee as requested by the superintendent.

5. *Physical and Mental Demands:*

   Physical:

   Maintain high level of physical stamina to supervise students throughout the day, including classes, labs, recesses, all-day field trips, special duties, night meetings, etc.

   Some light lifting (up to 40 pounds).

   Mental:

   Ability to communicate, both orally and in written form, in standard English.

   Ability to enforce safety standards in large group setting (e.g., lab safety).

   Ability to maintain emotional control under stress and to use judgment, patience, and sensitivity in dealing with students.

   Ability to respond to injuries and administer first aid as needed.

6. *Working Conditions:*

   Will have own classroom, but may have to share classroom space with other teachers (summer/night school).

   Daily work hours at school site, 7:30 a.m. to 3:45 p.m.; traditional school-year schedule, but must be willing to change to year-round schedule if required by district.

   Requires evening and weekend time commitments for special activities.

   Work environment may be physically isolated from coworkers (e.g., in portable classrooms).

   May have to cope with physically dangerous situations (e.g., handling of hazardous chemicals, potential violence in school and nearby neighborhood).

Figure 2.2. Continued

they offer documentary evidence about the nature of the job and the roles and responsibilities the job incumbent is expected to fulfill.

What Are the Next Steps in Developing a Teacher Evaluation System?

Once teachers' roles and responsibilities have been delineated, the next three steps are as follows:

- Develop the criteria for the evaluation and the indicators for those criteria.
- Determine which assessment methods and data sources to use.
- Set standards for evaluating performance.

Developing Criteria and Indicators

A criterion provides a general dimension against which performance, such as teaching, is rated or judged as meeting a standard or as successful or meritorious (Wheeler & Haertel, 1993). Indicators are the more specific and observable types of knowledge, skills, abilities, behaviors, and attributes that are empirically or by definition connected to the criterion. Descriptors are specific examples of the performance being assessed. They are useful for training people to implement the evaluation system and for informing teachers about the types of performance being assessed. However, they represent only examples of performance for an indicator and not the full scope of potential performance related to an indicator. Both indicators and descriptors must be stated specifically and in measurable or observable terms. Examples of some indicators and descriptors for given criteria are provided in Table 2.2.

Each indicator should focus on one aspect of teaching; that is, an indicator should not be multidimensional. An example of a multidimensional indicator—one that looks at two aspects of teaching performance—is "The teacher speaks clearly to the class and responds to students' questions in a timely manner."

TABLE 2.2 Examples of Criteria, With Indicators and Descriptors

Criterion: The teacher keeps parents informed about their children's academic progress and behavior in school.

Indicators	Descriptors
The teacher contacts each family by telephone or mail within the first month of school.	The teacher's log indicates that parents of seven children had not been contacted by the end of the first month of school, and that for five of the students, no effort had been made to contact the parents.
The teacher contacts all parents of Title I students at least once a month to inform them of their children's progress in the reading program and to monitor the students' home reading activities.	Notes were sent by the teacher to parents of six Title I students during the first week of November with a list of books they read in class in October, and copies were placed in the students' folders in the office.
The teacher notifies parents whenever their children receive awards for good behavior from the school staff.	A copy of the letter to the parents of the science fair winner is in the teacher's portfolio.

Criterion: The teacher maintains a physically safe environment with regard to equipment, materials and supplies, facilities, and numbers of people present.

Indicators	Descriptors
The teacher has the students using safety equipment (e.g., approved eye protection devices when working with hazardous materials and equipment and for designated athletic activities).	One parent wrote a letter to the principal complaining that the physical education teacher did not allow his daughter to ride with the school's Bike Club on October 4 because she did not have a suitable helmet. (School policy requires that all persons on school bike trips wear helmets that meet Snell standards.)

(continued)

TABLE 2.2 Continued

Chemicals are stored in accordance with compatibility groupings.	The teacher has arranged all chemicals in one collection by alphabetical order, not by compatibility groupings.
Chemicals are stored in appropriate, locked cabinets, off the floor.	The chemicals are stored in locked cabinets under lab counters designed for such use.
All chemicals are labeled, including preparation or receipt date.	Three bottles of chemicals on the third shelf of the second cabinet are not labeled.
The teacher spreads students out in the lab or workshop so they are not working in overcrowded or unsafe conditions.	There are 33 students in a physics lab of 1,200 square feet, considered overcrowding by the National Science Teachers Association's recommended standard of 45 square feet per student.
Emergency supplies, including a well-stocked first aid kit, a spill kit, and fire extinguisher, are readily available at all times.	Students report that their music teacher always has a supply of water bottles available on hot days before the marching band practices on the field.

Criterion: The teacher continually acquires updated and expanded knowledge of the subject area(s) he or she teaches or is eligible to teach.

Indicators	Descriptors
The teacher satisfactorily completes relevant courses in higher-education institutions.	The university transcript shows a grade of B+ in Russian history for the fall term.
The teacher attends school and district workshops and seminars.	The teacher's professional development log indicates that she went to two workshops on reading strategies and one on promoting students' writing skills.

TABLE 2.2 Continued

The teacher systematically reads professional literature and periodicals.	The teacher reports regularly reading two math journals, to which he subscribes.
The teacher participates in job-related studies, projects, and professional activities.	The teacher has been an essay reader for the state assessment program this year, participating in the reader training program in August.
The teacher goes to subject-matter conferences and meetings.	The teacher attended the state's English teachers conference in November this year.
The teacher writes papers and articles, conducts workshops, and makes presentations for colleagues in the same subject area.	The teacher did a workshop on integrating reading and writing instruction for the teaching staff at three schools in the district this summer.

The use of a solid foundation will help to ensure that valid criteria for evaluating teacher performance are delineated. The criteria should cover all the major areas of teaching (e.g., the five major categories of the DOTT; see Figure 2.1).

Determining Assessment Methods and Data Sources

Once the criteria and indicators are identified, the methods for assessing performance for each of them and the sources for data can be ascertained. Numerous assessment methods may be considered (Wheeler, 1994a); several possibilities are discussed in this book. There are also numerous possible sources of data, including educators and other people, existing records and data, teacher products, and other products (Wheeler, 1992b).

BEST COPY AVAILABLE

In some cases, the methods and data sources are readily apparent. For example, if teachers are to communicate with the parents or guardians of each student at least once a quarter, then two obvious data sources are teachers and parents. Assessment methods could include a review of the teacher records (phone logs and written communication file) and a survey of parents.

In other cases, the methods and data sources are not so readily apparent. For example, a teacher should be able to use the audiovisual equipment and computer technologies appropriate for his or her teaching assignment. To observe a teacher using all the possible equipment and technologies could be very time-consuming and would have to be done by a person knowledgeable about them. It would have to be determined which equipment and technologies are appropriate for a given teacher's assignment (both subject matter and types of students) and which are available for the teacher's use within the teaching context and work setting. Teachers could be required to maintain logs of their use of various equipment and technologies, or could be given knowledge tests about the equipment and technologies; students could be asked about their teachers' use of various types of equipment and forms of technologies; or various pieces of equipment and technologies could be chosen at random, and teachers asked to demonstrate their use and discuss their role in instruction.

Using the set of criteria derived from the list of roles and responsibilities and the indicators of performance, one can identify potential assessment methods and data sources. It should be kept in mind that, as Shulman (1988) points out in his discussion of several methods:

> Each of these several approaches to the assessment of teachers is, in itself, as fundamentally flawed as it is reasonably suitable, as perilously insufficient as it is peculiarly fitting. What we need, therefore, is a union of insufficiencies, a marriage of complements, in which the flaws of individual approaches to assessment are offset by the virtues of their fellows." (p. 38)

Wheeler (1994a) reminds us of two reasons for using multiple assessment methods: "(1) No one instrument or method is appropriate for all aspects of teaching performance covered by the evaluation system. (2) The use of multiple assessment methods for one domain or indicator allows for verification of data and triangulation of results" (p. 4).

Setting Standards for Evaluating Performance

Satisfactory levels of performance on criteria are specified by standards and are linked to the results on the assessments used to measure performance. Standards indicate what level of performance is considered minimally acceptable for a specified purpose and in a given teaching context. They should be set after the assessment methods are developed and prior to the implementation of the evaluation system. Procedures should be in place to ensure that the standards are reviewed regularly (at least every 5 years) and systematically to ensure that they remain appropriate, especially if the teaching context or school environment changes. For example, an influx of non-English-speaking students, a change from a standard school year to a year-round schedule, changes in staff configuration (e.g., eliminating all counselors, adding a reading specialist), or the acquisition of several computers for each classroom could affect the appropriateness of current standards.

There are several types of standards (Wheeler & Haertel, 1993). *Developmental* standards are often used for professional improvement, mentoring, training, self-evaluation, and formative evaluation. They indicate the level of growth or change to be attained by teachers, individually or as a group, and apply to new as well as more experienced teachers. *Minimum* standards designate the level of performance that must be met to satisfy some requirement, and below which performance is not acceptable. Such standards are used for licensure and credentialing, hiring, and personnel decisions such as job assignments and retention or termination. There may be different minimum standards for new and experienced teachers, or for teachers working in different contexts (e.g., at schools in low-income areas, with students who have physical disabilities, at schools with no counselors). *Desired* performance standards reflect exemplary or meritorious levels of performance. These types of standards apply to such uses as promotions, awards, and special recognition. Examples of these types of standards are shown in Table 2.3.

TABLE 2.3 Examples of Different Types of Standards for Three Criteria (with possible assessment methods)

Criterion: The teacher gives appropriate assignments, in and out of class, to students.

Developmental: The teacher becomes better able to modify assignments in the teacher's guides for non-English-proficient students as shown by progressively better adaptations of five different assignments each, over a 3-month period, in language arts, mathematics, and science. (portfolio)

Minimum for new teacher: The teacher selects appropriate classroom activities and homework assignments from the teacher's guide or instructional kit for use with his or her own students. Appropriateness considers students' needs and abilities, instructional program goals and objectives, and available resources. (observation, portfolio)

Minimum for experienced teacher: The teacher selects and modifies appropriate classroom activities and homework assignments from the teacher's guide or instructional kit, and develops new ones for use with his or her own students. Appropriateness considers students' needs and abilities, instructional program goals and objectives, and available resources. (observation, portfolio)

Desired performance: The teacher develops new, innovative, and appropriate classroom activities and homework assignments for use with his or her own students and by other teachers. Appropriateness considers students' needs and abilities, instructional program goals and objectives, and available resources. (observation, portfolio)

Criterion: The teacher manages student misbehavior in the classroom and elsewhere at school.

Developmental: The teacher acquires skills in anger management and conflict resolution by completing this year's training program. (professional development log)

Minimum for new teacher: The teacher maintains enough control of his or her students so that no student is physically injured or emotionally abused. (observation, office records)

TABLE 2.3 Continued

Minimum for experienced teacher: The teacher maintains enough control of his or her students so that there are few and insignificant disruptions to learning and other activities. (observation, office records)

Desired performance: The teacher is able to handle quickly any potential or actual violent situations on the school campus. (peer reports, observation, simulated work sample)

Criterion: The teacher adapts instructional content and concepts to address students' understanding of the subject matter and their misconceptions.

Developmental (for a more experienced teacher): The teacher plans and conducts a workshop for the teaching staff to impart expertise that will help them address common student misconceptions about weather and climate. (peer feedback, science test of student understanding)

Minimum for new teacher: The teacher assumes that students will understand the content and concepts being presented in class, and explains the material in the same manner to all students. (observation)

Minimum for experienced teacher: The teacher realizes during instruction that some students are confused by the content and concepts presented in class, and explains the materials in more than one way to improve student understanding. (observation)

Desired performance: The teacher anticipates in advance what misconceptions the students will have and what content and concepts might be confusing to them, and prepares explanations prior to instructing the students. (observation, portfolio)

An array of methods, both empirical and judgmental, exist for setting standards (Cizek, 1996; Livingston & Zieky, 1982). Most methods employed to set standards for evaluating teaching performance rely heavily on professional judgment. The process of establishing standards should involve key stakeholders—teachers, administrators, school board representatives, and others. These individuals must be carefully selected and thoroughly trained. They should be familiar with and well-informed about the array of teaching positions and assignments to be covered, the roles and responsibilities associated with those positions, the teaching context, the criteria and indicators to be covered, the nature of the assessment methods, how the assessments will be scored, and who will use the evaluation results and for what purposes. Cizek (1995) warns of the importance of training those involved in setting standards:

> [An] area in desperate need of attention is the training provided to standard setting participants. In the legal arena, it can take months to select a jury and to educate them about the task they will be asked to perform. Indeed, in other areas such as athletics, music, or drama, the vast majority of time is spent in training for but a fleeting moment of performance. In contrast, the training provided to standard setting participants is often minimal compared to the task they will be asked to perform. (pp. 10-11)

Whatever methods are used, the standards should not be set so low that they have the unintended consequence of dragging down the quality of teaching. They must be fair, reasonable, attainable, and clear to all persons involved in the evaluation process, including the teacher. They must be designed to encourage good teaching, promote student learning, and improve student behavior. Above all, they must protect the interests and well-being of the students, their families, the school, and the community.

Summary

In the selection of a foundation on which to build a teacher evaluation system, the first question that must be asked is, What is the purpose of the evaluation system? Some of the foundations discussed in this chapter are much more appropriate than others, or may even be required, for some purposes, as in the case of state laws and government regulations. Other factors that must be considered include (a) the availability of the information and of the people needed to use a particular foundation (e.g., local or applicable job analysis data, well-constructed lists of teacher duties and job descriptions, someone to conduct a review of the research literature), (b) the quality of this information, and (c) the qualifications of people who would provide input and judgment. More than one foundation will have to be employed for one or another uses, even if the evaluation system builders rely heavily on one foundation for most purposes.

The use of a combination of foundations results in a more comprehensive and solid basis for a multipurpose teacher evaluation system than does the use of any single foundation alone, no matter how strong that foundation is. This will lead to teacher evaluation systems that permit better decisions to be made and that will be more apt to improve the overall quality of a wider array of teaching or teaching-related performances. It can also strengthen the technical, professional, and legal quality of the system.

Through the use of appropriate foundations, valid criteria and indicators of teaching performance can be derived. Based on the performance indicators, suitable and feasible methods of assessment can be selected or developed, and sources of data can be identified. Standards for minimally acceptable performance for each criterion can be set. These standards should reflect the purpose of the teacher evaluation system, incorporate the assessment methods, and protect the interests and well-being of all concerned parties, especially the students.

References

American Association of School Personnel Administrators. (1995, April). *Most critical knowledge and skills of a future educator*. Sacramento, CA: Author.

Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education, 1*, 345-363.

Braskamp, L. A., Brandenburg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills, CA: Sage.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York: Macmillan.

Centra, J. A. (1980). *Determining faculty performance*. San Francisco: Jossey-Bass.

Cizek, G. J. (1995, April). *Standard setting as psychometric due process: Going a little further down an uncertain road*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Cizek, G. J. (1996). Setting passing scores (NCME Instructional Module). *Educational Measurement: Issues and Practice, 15*(2), 20-31.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought process. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York: Macmillan.

Costa, A. L., Garmston, R. J., & Lambert, L. (1988). Evaluation of teaching: The cognitive development view. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 145-172). Alexandria, VA: Association for Supervision and Curriculum Development.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. *Educational Evaluation and Policy Analysis, 8,* 45-60.

Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to assess systems of evaluating educators.* Newbury Park, CA: Sage.

Joyce, B., & Weil, M. (1986). *Models of teaching* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores.* Princeton, NJ: Educational Testing Service.

McConney, A., Wheeler, P. H., Wiersma, W., Millman, J., Stufflebeam, D., Gullickson, A., Airasian, P. W., Vicinanza, N., Haertel, G. D., Barry, K., & Cullen, K. (1996). *Teacher evaluation kit and database of CREATE products* [CD-ROM]. Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

McDonald, F. J., & Elias, P. (1976). *The effects of teaching performance on pupil learning* (Final Report, Beginning Teacher Evaluation Study, Phase II, Vol. 1). Princeton, NJ: Educational Testing Service.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach.* New York: Longman.

Millman, J. (1981). Student achievement as a measure of teacher competence. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 146-166). Beverly Hills, CA: Sage.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

Ornstein, A. C. (1991). Teacher effectiveness research: Theoretical considerations. In H. C. Waxman & H. J. Walberg (Eds.), *Effective teaching: Current research* (pp. 63-80). Berkeley, CA: McCutchan.

Robinson, G. E. (1984). *Incentive pay for teachers: An analysis of approaches*. Arlington, VA: Educational Research Service.

Rosenfeld, M., Thornton, R. F., & Skurnik, L. S. (1986). *Analysis of the professional functions of teachers: Relationships between job functions and the NTE Core Battery* (Research Report 86-8). Princeton, NJ: Educational Testing Service.

Scriven, M. (1988a). Duty-based teacher evaluation. *Journal of Personnel Evaluation in Education, 1*, 319-334.

Scriven, M. (1988b). Evaluating teachers as professionals: The duties-based approach. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 110-142). Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1990). Can research-based teacher evaluation be saved? *Journal of Personnel Evaluation in Education, 4*, 19-32.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Scriven, M. (1994a). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8*, 151-184.

Scriven, M. (1994b). Using the duties-based approach to teacher evaluation. In L. Ingvarson & R. Chadbourne (Eds.), *Valuing teachers' work: New directions in teacher appraisal* (pp. 70-95). Melbourne: Australian Council for Educational Research.

Shuell, T. J. (1993). Toward an integrated theory of teaching and learning. *Educational Psychologist, 28*, 291-311.

Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*(3), 36-41. (ERIC Document Reproduction Service No. EJ 385 344)

Wheeler, P. H. (1991, November). *Building a teacher evaluation system: Finding the right foundation.* Paper presented at the annual meeting of the California Educational Research Association, San Diego.

Wheeler, P. H. (1992a). Foundations for building teacher evaluation systems. *CREATE Newsletter, 2*(2), 4-5.

Wheeler, P. H. (1992b, October). *Sources of data for evaluating teachers* (TEMP B Memo 7). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Wheeler, P. H. (1994a, August). *Assessment methods for use in evaluating educators* (TEMP C Memo 12). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Wheeler, P. H., with Scriven, M. (1994b). Duties of the teacher: An essential foundation for a teacher evaluation system. *AASPA Report, 1*(7), 2-3.

Wheeler, P. H. (1994c). *Foundations upon which to build a teacher evaluation system* (TEMP C Memo 18). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Wheeler, P. H. (1995, July 10). *Building and rebuilding teacher evaluation systems: Selecting the appropriate foundation.* Paper presented at the fourth annual National Evaluation Institute, Kalamazoo, MI. (Also available as EREAPA Publication Series 95-1, EREAPA Associates, Livermore, CA)

Wheeler, P. H. (1995-1996). Consider these issues before you use student tests in teacher evaluation. *AASPA Report,* 3(3), 12-13.

Wheeler, P. H., & Haertel, G. D. (1993). *Resource handbook on performance assessment and measurement: A tool for students, practitioners, and policymakers.* Chico, CA: Owl. (ERIC Document Reproduction Service No. ED 367 686)

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices.* Santa Monica, CA: RAND Corporation.

3

Legal Considerations in Designing Teacher Evaluation Systems

PAMELA D. TUCKER
KAY P. KINDRED

The legal context for teacher evaluation has its underpinnings in constitutional, statutory, and case law, but fundamentally, these laws strive to balance the legal rights of the individual against the rights and needs of society. The rights of the individual teacher to due process are balanced against the obligation of the school board to ensure the quality of instruction that students receive. Although the courts are often viewed as favoring the rights of teachers over those of school boards, which act on behalf of students, analysis of dismissal cases has shown that school districts have prevailed in the majority of cases (Cain, 1987; Rossow & Parkinson, 1992). Typical of many legal decisions, the court in *Conley v. Board of Education* (1956) commented that "a school board [has] the right to demand that a teacher know his subject and that he be capable of arousing and holding the interest of his pupils and maintaining discipline" (cited in Tigges, 1965, p. 1095).

Despite a record of support by the courts in cases where there has been proper due process, school administrators are often leery of the legal ramifications of teacher evaluation in cases where there is potential for nonrenewal or dismissal (Cain, 1987; Fuhr, 1993; Laing, 1986; McGrath, 1993). Time, effort, and money become important considerations when decisions must be made concerning whether to identify and work with teachers who are performing less than satisfactorily. However, the courts, as well as researchers in the field of teacher evaluation, have observed that school districts have a *responsibility* to "hold teachers accountable to standards of practice that compel them to make appropriate instructional decisions on behalf of their students" (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984, p. 80).

Our purpose in this chapter is to define the legal context and parameters for designing and conducting teacher evaluation in a reasonable and fair manner. We begin with an overview of the purposes of teacher evaluation and the sources of legal standards for evaluation, which is followed by discussion of statutory and case law examples of the desirable elements of evaluation procedures and processes. The discussion focuses primarily on the practical features of an evaluation system and procedures for implementation. Finally, we recommend some principles for designing and implementing an evaluation system, drawn from various legal sources.

Why Evaluate?

The centrality and importance of teacher effectiveness in the ongoing pursuit of quality education are unquestionable. Although curricula and other variables play an important part in the educational experience, instructional expertise is at the heart of the learning enterprise (Corcoran & Goertz, 1995; Crone & Teddlie, 1995; Shapiro, 1995; Wang, Haertel, & Walberg, 1993). The ability to judge teacher effectiveness accurately is essential to all the personnel processes (e.g., recruitment, promotion), but most especially to the ongoing evaluation of classroom teachers, who ultimately determine the quality of the whole educational process (Iwanicki, 1990). Personnel evaluation, in general, and teacher evaluation, in particular, are typically viewed as serving the dual purposes of accountability (i.e., decision making regarding teacher effectiveness) and professional development. For administrators, personnel evaluation can be one of the primary means of ensuring a quality educational program for students, and yet many are hesitant to conduct honest and meaningful evaluations with staff for fear of the potential legal ramifications in cases of unsatisfactory performance (McGrath, 1993).

The purposes of teacher evaluation are conceptualized variously by different authors, but Wise et al. (1984, p. 11) offer one characterization that seems relevant to actual practice in schools. The purposes are both accountability and improvement at the individual and school levels, specifically the following:

1. Individual personnel decisions, to ensure that only effective teachers continue in the classroom
2. Individual staff development, to foster the professional growth of new and continuing teachers
3. School status decisions, to provide a basis for judgments such as accreditation
4. School improvement, to guide the school in responding to student needs

Other authors have categorized the purposes as primarily formative and summative (Millman, 1981; Scriven, 1967). Most legal references on personnel evaluation use *formative* to indicate the developmental process of collecting and sharing information on the teacher's performance and *summative* to indicate the final synthesis of the documentary material for the personnel file; these are treated as distinct stages and components of the evaluation process (Beckham, 1985a).

Teacher evaluation currently relies on three primary modes of assessment: teacher competency tests, teacher rating scales, and measures of teacher effectiveness based on learner gains (Alkin, 1992). Initial certification of new teachers entails the use of both rating scales during the internship and teacher competency tests for entry into the profession (Alkin, 1992). There is also a growing trend to use competency tests, such as the National Teacher Examination (NTE), for recertification requirements (Rossow & Parkinson, 1992). Constitutional challenges have been raised regarding the potential of such tests for discrimination based on their disparate impact on minorities. Other legal problems involve their job-relatedness and validity.

Once a teacher is hired, rating scales are often used in combination with observation and are the most common means of determining contract renewal and eventual tenure. In an Educational Research Service (ERS; 1988) study of teacher evaluation in 909 school districts, 99.8% of the responding school districts reported using direct observation as part of the evaluation process. In addition to observation, 87.4% of the school districts used checklists of instructional and professional behaviors to evaluate teachers. Despite strong opposition by teachers' unions, 67% of the school systems surveyed by the ERS were found to use some measure of learner gains. As Alkin (1992) has noted, the "idea of basing decisions about teachers on measures of how much students learn from them is one that the public and its elected representatives find very appealing" (p. 1349), but it is also considered problematic given the vast number of variables affecting student learning. Alkin observes that, despite their wide use, "none of the three principal strategies available for evaluating teachers (tests, rating scales, and student learning) works satisfactorily, so there is ample room for improvement" (p. 1345).

One of the new trends in teacher evaluation that addresses concerns about the traditional reliance on classroom observation is the use of multiple sources of information on teacher performance or multifaceted data collection. Stronge and Helm (1991) identify five potential categories of data sources for assessing personnel performance: "observation, self-assessment, questioning, artifact collection, and documentation" (p. 60). Questioning, as defined by Stronge and Helm, includes peer evaluations, student ratings, and parent ratings. These authors assert that "the most complete picture of an employee's performance will be obtained by questioning at least a representative sampling from all of the various constituencies with whom the person works" (p. 180). Despite support for the use of multiple data sources, research on actual practice has found that peer evaluations are used in less than 6% of the schools, student ratings in less than 3%, and parent ratings in less than 1% (ERS, 1988).

As the most visible professional within the school environment, the classroom teacher has been evaluated in some manner for as long as we have had schools. Over the years, evaluation has evolved into the highly structured and prescribed procedures used in most school systems today. Much of the current prescriptiveness is a result of public attention and legislation passed during the 1970s and 1980s (Fuhrman, 1994). Beginning in the 1970s with a push for accountability, the general public demanded more rigorous performance assessments of school programs and personnel to demonstrate educational effectiveness (Harris, 1981). The pressure to improve the quality of teaching was heightened by reports such as *A Nation at Risk* (National Commission on Excellence in Education, 1983), and many state legislatures began to mandate evaluation not only of teachers but also of all certificated employees; subsequently, evaluation became the primary tool of accountability in the schools (McLaughlin & Pfeifer, 1988). This expansion of the "state role in defining and developing evaluation systems for teachers" (Sclan, 1994, p. 2) resulted in the implementation of teacher evaluation systems by most states.

What Are the Sources of Legal Requirements in Teacher Evaluation?

Teacher evaluation is a process conducted within and constrained by the application of legal principles. It is the consideration of those principles that dictates the framework for development of evaluation criteria and procedures and determines the requisite accommodations applicable in cases of unsatisfactory performance. The legal parameters of teacher evaluation derive from several sources: federal constitutional law and state statutory law, and the judicial decisions that interpret and apply them. The outcome in any particular case will, of course, turn on the specific statutes, policies, negotiated agreements, and circumstances involved; thus, differences may exist in what constitutes an acceptable evaluation procedure from state to state and from circuit to circuit (Frels, Cooper, & Reagan, 1984). Nonetheless, it is possible to define the structural underpinnings common to a legally sound system of public teacher evaluation.

Constitutional Foundation

The due process clause of the Fourteenth Amendment of the U.S. Constitution is the foundation for the guiding principle of fairness in the evaluation process. The due process clause prohibits government action that would deprive individuals of "life, liberty, or property, without due process of law." *Due process* refers to both the *process* of decision making (procedural due process) and the *product* of that decision making (substantive due process). The substantive component of due process may protect against the violation of certain fundamental rights or may void arbitrary restrictions on individual freedom of action. The procedural aspects of due process ensure that when the power of government is to be used against an individual, a fair procedure will determine the basis for, and the legality of, that action. But only when the deprivation of an individual's "life, liberty, or property" is at issue does the government owe that person some type of process for the consideration of his or her interests (Nowak & Rotunda, 1995).

When the government acts as employer, there are special issues with respect to the existence of liberty or property rights in employment. As state agents, public schools must provide due process when an employee's liberty or property rights might be denied. However, "due process rights of teachers do not shield them from termination. Incapable or insubordinate teachers can and should be terminated" (Frels & Horton, 1994, p. 1).

In *Board of Regents v. Roth* (1972), the U.S. Supreme Court clarified the meaning of liberty and property rights in the context of nontenured teaching positions. Roth was an assistant professor hired for a 1-year term at Wisconsin State University at Oshkosh. He had no tenure rights, and under state law the decision whether to rehire a nontenured teacher was left fully to the discretion of university officials. The Court stated that a nontenured teacher has a liberty interest not to be stigmatized or to have anything be said that damages his or her good name or reputation. Thus, the violation of a liberty interest requires some evidence that the teacher's "good name, reputation, honor, or integrity" had been negatively affected or that the personnel decision had "imposed on him a stigma or other disability that foreclosed his freedom to take advantage of other employment opportunities" (pp. 572-573). Although this is useful as a general standard, the Court did not specifically define *stigma* (Yudof, Kirp, & Levin, 1992). In Roth's case, the Court held there had been no deprivation of liberty.

In sum, mere dismissal from a particular position does not implicate a liberty interest because having one form of government employment foreclosed does not constitute the kind of deprivation of freedom encompassed by the term *liberty*. Only if, in dismissing the teacher, the government also forecloses his or her chances of employment in a wide range of activities might the dismissal encompass a deprivation of liberty sufficient to require that the individual receive a due process hearing (Nowak & Rotunda, 1995). Because personnel matters are typically considered confidential, teachers are rarely able to establish a violation of liberty interests.

The violation of property rights is a more viable claim in teacher dismissal cases, especially when the teacher has a continuing contract. In *Roth*, the Court established that for a property right to exist, one must have a "legitimate claim of entitlement" (p. 577) to some benefit, and in its companion case, *Perry v. Sindermann* (1972), the Court explained that there must be some "mutually explicit understanding" (p. 601) of the benefit. Thus, a person will have a constitutionally recognizable property right in a government benefit, such as public employment, if the person can be deemed to be "entitled" to the benefit. The applicable federal, state, or local law that creates the benefit must define the interest in such a way that there is a mutual expectation that the individual should continue to receive the benefit under the terms of the law in order for there to be a recognized claim of entitlement (Nowak & Rotunda, 1995). Based on these standards, Roth had no property interest because he had only a 1-year contract, whereas Sindermann, who, though not tenured, was in fact entitled by Texas state law to continuing reemployment absent nonrenewal for cause, was found to have an implied property right based on administrative policies and practices.

State statutes often speak to the issue of whether a property interest is granted by the probationary and continuing contract. Case law has further clarified that a tenured teacher's teaching position and salary increments do constitute property rights, but additional duties such as extracurricular work and textbook selection are held at the will and pleasure of the school board and are not constitutionally protectable interests (*Needleman v. Bohlen*, 1979). In addition to establishing a denial of either liberty or property rights, the teacher must also establish that the deficiencies that were given as grounds for dismissal were "arbitrary and capricious and not supported by credible evidence" (Delon, 1982, p. 62).

Finally, the equal protection clause of the Fourteenth Amendment demands that all employees be treated equally and subject to the same provisions of the law. Legal challenges have been based on policies that unconstitutionally classify teachers, such as the use of standardized tests (e.g., the National Teacher Examination). However, few challenges have been successful because "the state action is presumed to be valid" and "rationally related to a legitimate government interest" (Rossow & Parkinson, 1992, p. 54) by the courts unless proven otherwise.

Numerous court cases have addressed adverse personnel decisions that involved the actual or alleged violation of First Amendment rights. Typically these concerns are raised when teachers have been active in teacher associations or unions. Under these circumstances, teachers are required to present factual information that supports their allegations (*Montgomery v. Trinity Independent School District*, 1987), and if they are successful, schools must then prove that the same actions would have been taken regardless of a constitutionally protected behavior such as free speech or association (Beckham, 1985a).

Statutory Law

By 1950, all states had enacted statutes addressing the dismissal of incompetent teachers (Adams, 1988/1989). The grounds for dismissal vary from state to state, ranging from states that use a single criterion of very general language to those with multiple causes for dismissal, such as Nevada, with 16. Although all the statutory law is based on the Tenth and Fourteenth Amendments, the particular history of each statute and its interpretation by the courts has had a major influence on the resulting case law, so that substantial variations exist across states.

In addition to the state's involvement in defining the broad statutory terrain for dismissal, state legislatures also provide guidance on the specifics of personnel evaluation. As a result of the public demand for quality instruction, discussed earlier, and the courts' growing reliance on school districts' evaluation policies, state legislatures increasingly have played a role in defining the broad parameters for personnel evaluation based on local policies and procedures (Frels et al., 1984). A national survey conducted by Stronge and Tucker (1995) indicates that 79% of the responding states legally mandate evaluation of school personnel by state law, state board of education policy, or state superintendent directive. In some states, a combination of mandates exists. In Hawaii, Texas, and New York, for example, all three sources of legal requirements for personnel evaluation are in place.

State statutory law is the most common source of mandatory teacher evaluation. Of the states responding to the Stronge and Tucker survey, 69% cited statutory requirements; 52% reported a supplemental or independent state board of education policy. Only 12% of the states reported the use of state superintendent directive regarding personnel evaluation. In all of these states, except New York, a superintendent's directive existed in addition to statutes or board policies. Of the 42 responding states, 8 (19%) indicated they had no state-level legal mandate for personnel evaluation.

In addition to legislating evaluation procedures, one state has taken an even bolder step, tying satisfactory evaluations to continuing teacher certification. The provisions of the Children First Act in Louisiana set forth that teaching certifications can be revoked based on unsatisfactory evaluations. When this law's constitutionality was challenged in *Eiche v. Louisiana Board of Elementary and Secondary Education* (1991), the Supreme Court of Louisiana ruled that provisions of the act relating to revocation of teaching certificates based on the teachers' failure to obtain satisfactory evaluations are not in violation of the state constitution.

State statutes typically define two levels of teachers—those on probationary contracts and those with continuing contracts. Because the due process rights afforded teachers differ based on the level of commitment that the state has made to them for continuing employment, the resulting due process procedures also differ based on the designation of probationary or continuing contract. Nonrenewal of a probationary teacher's contract typically requires notification that a contract will not be offered for the upcoming year, whereas dismissal of a teacher who is on continuing contract requires full due process, including a statement of specified deficiencies, an opportunity to improve, and an opportunity to question and respond to the reasons for dismissal (Alexander & Alexander, 1985). An unusual provision of the Texas law allows school districts to choose between offering teachers continuing or fixed-term contracts. Case law in Texas has further clarified that teachers have no implied property interest in continued employment under the fixed-term contract system (*Texas Education Code Annotated*, 1991/1995). In most states, the general evaluation process and criteria tend to be similar for both probationary and continuing contract teachers, with the exception of how often the evaluations are conducted.

In contrast to Texas, with its weaker protection for teachers with continuing contracts, Ohio offers considerable protection for teachers (Petrie & Black, 1983). The standard for termination in the *Ohio Revised Code* is that of "gross incompetency," which has come to be interpreted as "extreme, flagrant, or complete" inefficiency by the teacher (p. 1050) in the Ohio case law. This standard has placed a substantial burden of proof on local school districts compared with the usual standard of "incompetency" and has resulted in some daunting termination cases. "The most extreme case of attenuated teacher termination proceedings on record" (Petrie & Black, 1983, p. 1041) dragged on in the courts for more than 10 years (*Jones v. Mt. Healthy Board of Education*, 1983). Unfortunately, the *Ohio Revised Code* provides little guidance to the local school system on how to conduct teacher evaluation, although case law has helped to define what constitutes "gross inefficiency." A comparison of the statutorily defined grounds for dismissal in the states of Texas, Ohio, and California is presented in Table 3.2, later in this chapter.

Table 3.1 provides a checklist of possible due process safeguards in an evaluation system. The Xs in the three state columns in the table indicate the explicit presence of given items in the state codes. The omission of any of these items does not suggest that these due process options are not available, but merely that they are not addressed specifically in the codes. The *Texas Education Code* has been revised most recently and is the most explicit and prescriptive in terms of (a) the evaluation process, (b) grounds for dismissal, and (c) due process guarantees.

Case Law

We use case law that addresses the issues of teacher evaluation throughout this chapter to provide examples of how the intersection of state laws, negotiated agreements, and local school board policies has been interpreted by the courts. Due to the myriad possible combinations, each decision is unique to its specific circumstances. However, the interpretations from various courts across the country are somewhat consistent in the basic expectations concerning the teacher evaluation process. When the basic elements of due process

TABLE 3.1 Evaluation Components of State Statutes

	California	*Ohio*	*Texas*
State guidelines	X		
Frequency of evaluation	annually probationary; 1 to 2 times a year permanent	twice a year if intention is not to reemploy	annually
Assessment categories	X	X	X
Use of standardized tests	X		
Written evaluation	X	X	X
Opportunity to respond	X	X	
Conference	X	X	
Notice of unsatisfactory performance	X	X	X
Specific recommendations for remediation	X	X	X
Assistance	X	X	
Hearing	X	X	
Training for evaluators	X		
Special features	explicitly excludes use of exams	uses gross inefficiency as standard for dismissal	rating scale prescribed in statutes, one probationary year if conditions met in § 13.306

noted in Table 3.1 have been present, the courts have typically ruled in the school district's favor (Adams, 1988/1989; Cain, 1987; Rossow & Parkinson, 1992). Illustrative cases supporting the use of each evaluation component are provided in the pages that follow.

Negotiated Agreements

Teacher unions or associations have negotiated agreements that affect the employment of teachers in most states. Negotiated agreements typically address conditions of employment such as salary, class size, preparation time, and instructional time. The evaluation procedures have also been interpreted to affect the work and welfare of employees, and thus the state courts have permitted the negotiation of procedural aspects of the evaluation (e.g., *Wethersfield Board of Education v. Connecticut State Board of Labor Relations*, 1986). The specific evaluation criteria have been considered by most courts to be a management right or prerogative (e.g., *Ohio Revised Code*), and they are not a bargaining item (McCarthy & Cambron-McCabe, 1992).

In addition to providing procedural protections in the evaluation process, negotiated agreements can play a substantial role in the dismissal process by making it subject to grievance procedures (Johnson, 1984). In the case of a negative evaluation, an unsatisfactory rating, or a recommendation for dismissal, a teacher can file a grievance. During the grievance procedure, the teacher is provided with union representation at a structured hearing or series of hearings to review the procedures used in the evaluation process and sometimes the actual contents of the evaluation document (Frels et al., 1984). "Typically, as the appeal moves up the steps of the grievance procedure, the hearings become more formal and more adversarial" (Johnson, 1984, p. 121). Such a case represents a demanding challenge for administrators, who bear the burden of proof.

What Are the Recommended Components of an Evaluation System?

Because the combination of mandates varies by school district, there is no one evaluation system that conforms to the unique circumstances existing in all localities (Wise et al., 1984), but some general principles are fairly consistent. Evaluation procedures must

1. be clearly articulated and uniformly applied to meet the judicial standard of "reasonableness" and "fairness" (Beckham, 1985a),
2. address both the general substantive and procedural aspects of an evaluation system, and
3. provide "notice" to teachers of the conditions of employment when clearly articulated, explained, and distributed to every teacher.

To satisfy these basic principles, evaluation systems typically include the following components:

- A statement of purposes
- Performance criteria
- A rating scale that defines standards of performance
- A description of the procedures used to collect information on performance
- A formal means of summarizing the information on performance, such as an evaluation summary

A handbook describing the entire evaluation system, inclusive of the above items, should be distributed to all teachers before implementation (Frels et al., 1984). Such a handbook should provide notice of both the expectations and possible rewards or disciplinary action for performance.

Statement of Purpose

Evaluation policies and procedures *must* conform to statutory mandates, state and local school board regulations, and negotiated contracts. All policies of a local school board, including the evaluation policy, have the force of law, and the board must comply with its own rules (*Iverson v. Wall Board of Education*, 1994). Because the evaluation policy has the force of law, it should be written with care and should reflect principles that the board is capable of fulfilling. An evaluation policy should include a clear statement of purpose.

The statement of purpose in an evaluation policy can set the tone for the school system's approach to evaluation, establish its rational relationship to the state's purpose, and stipulate the possible consequences of the evaluation process (Delon, 1982; Frels et al., 1984). Typically, educational authors have emphasized the dual purposes of teacher evaluation as accountability and professional development (Darling-Hammond, 1990; Iwanicki, 1990; Millman, 1981; Stronge & Helm, 1991). Translated into practice, these purposes can include tenure decisions, promotion, merit pay increases, and termination. A clear statement of the tangible consequences of the evaluation process provides employees with "notice" of the potential state action, an important element of due process. In addition to delineating the uses of the evaluation process, the policy should state when it is to take effect and whether it supersedes in part or whole a previous policy (Frels et al., 1984).

Criteria

The most critical issue in the assessment process is the development of the criteria against which a teacher's performance will be measured (Frels et al., 1984). There should be "sufficient specificity in the elaboration of assessment standards so as to inform a reasonably prudent person of the applicable criteria" (Beckham, 1985a, p. 9). Case law and most state statutes require that performance criteria be objective and job related (Beckham, 1985a). The courts, however, have acknowledged and accepted that there is a subjective quality to evaluation that is unavoidable. For example, in *Rogers v. Department of Defense* (1987), the court found that although the evaluation required some subjective judgment, the overall process (including meetings and written instructions) was sufficiently objective to inform the teacher of areas in need of improvement. The subjectiveness of the judgments is permissible so long as the criteria are job related, observable, and uniformly applied to all teachers (Beckham, 1985a).

In *Eshom v. Board of Education of School District No. 54* (1985), a tenured teacher argued that the principal's evaluation of her teaching was invalid because he compared her with other teachers in the building instead of to objective standards. The Supreme Court of Nebraska ruled that teacher performance could not be "measured in a vacuum nor against a standard of perfection but, instead, must be measured against the standard required of others performing the same or similar duties" (p. 13). Therefore, performance criteria must be applied in a manner that balances an absolute and normative interpretation of their meaning. This process is, of course, less than objective, but it is legally permissible by the courts if there is uniform application.

If some criteria are considered more important than others, this should be indicated in the written evaluation manual (Beckham, 1985b). The relative importance of various criteria was also addressed in the *Rogers v. Department of Defense* (1987) case. The school stipulated that two of the five job elements were "critical for satisfactory job performance," specifically that of "student assessment" and "instructional program." The teacher claimed that he was rated unsatisfactory on only a few of the subitems in each of these categories and that his overall performance did not warrant an unsatisfactory rating. The court held that the school could satisfy its burden of proof regarding an overall unsatisfactory rating based on fewer than all of the subitems of a critical job element if (a) the teacher was given adequate notice of the importance of these elements or (b) evidence was provided that supported the critical nature of unsatisfactory subitems in the overall job performance. In this case, both conditions were met and the teacher's dismissal was upheld.

Specific performance standards are typically left to the discretion of the local school board, with minimal guidance from state law (Rossow & Parkinson, 1992). For example, the *California Education Code* (§ 44662) states that local school districts should evaluate certificated employees in the following areas: (a) progress of students toward established standards, (b) instructional techniques and strategies, (c) adherence to curricular objectives, and (d) establishment and maintenance of a suitable learning environment (*California Education Codes Annotated*, 1993a). In addition to performance standards that address specific classroom duties, including disciplinary practices (*Sargent v. Selah School District No. 119*, 1979), criteria can include items that assess responsibilities such as coaching, supervision of yearbook development, and professional conduct. If performance criteria can be justified as job related and a nexus, or connection, exists between the identified behavior to be evaluated and proper fulfillment of the professional role, they are generally acceptable to the courts.

California's inclusion of the criterion of student progress represents a relatively new and controversial development in teacher evaluation. Although it is assumed that student learning has been implicit in the educational enterprise since the beginning, explicit inclusion of some measure of student achievement is a more recent practice. Researchers in the area of educational evaluation and teachers' unions have criticized the practice because they argue that student progress cannot be measured in a vacuum. Student achievement has been found to be influenced by individual differences in ability and a multiplicity of other variables that are often beyond the control and influence of classroom teachers (Alkin, 1992). Nonetheless, several states include provisions similar to California's, and a number of court decisions, using the rational basis standard, have upheld the use of tests to assess student progress as a reasonable basis for unsatisfactory evaluations (Petrie & Black, 1983; Rossow & Parkinson, 1992).[1]

A second controversial area in the selection of performance criteria for teacher evaluation is the use of standardized testing results for teachers. Opinions vary on this issue, as exemplified by California statutes that explicitly preclude the use of teacher testing (*California Annotated Code*, 1993a, § 44662) and Texas statutes that require it (*Texas Education Code Annotated*, 1991/1995, § 13.047). In *United States v. South Carolina* (1978), the Supreme Court upheld the use of the National Teacher Examination as permissible to establish a minimal level of knowledge within a subject area, but noted that the examination does not measure teaching skills or teaching effectiveness. Therefore, although valid tests such as the NTE may be used for assessing content-area knowledge, they have not been accepted as indicative of teaching performance (Adams, 1988/1989).

Rating Scale

"A rating system must have definition and meaning to all [parties] involved in the assessment process" (Frels et al., 1984, p. 5). The ratings given to a teacher should be consistent with any directives for improvement. In *Long v. School District of University City* (1989), a teacher received an overall rating of "good" but was informed that her performance in specific areas needed to improve or she would not receive a raise the following year. According to the administrator, the teacher did not improve in the specified areas the following year, was given an unsatisfactory rating, and subsequently did not receive a salary raise. When she requested a hearing before the school board, her request was denied. When the teacher took the case to court, a lower court ruled that she had been denied due process rights, especially in the form of a hearing, and the Missouri Court of Appeals upheld the decision. In this case, the incongruity between the overall rating of "good" one year and the consequence of a salary freeze the following year raised the question of arbitrary and capricious action.

Rating scales or systems are usually developed at the local level, but in some states, such as Texas, rating scales are prescribed by state statute (*Texas Education Code Annotated*, 1991/1995, § 13.306). In either case, the more precise the definition for each rating, the more objective and credible the rating scales generally will be perceived to be by teachers and the courts.

Procedures

A description of evaluation procedures should identify who will be evaluated, by whom, and how often. According to the 1988 Educational Research Service survey cited previously, 99.6% of the 909 reporting school districts had formal evaluation procedures for probationary teachers, and 98.7% had such procedures for tenured teachers. However, the quality of the procedures and their utility in achieving the stated purposes vary substantially (Sclan, 1994). Even the frequency of evaluations varies based on the district and on the longevity of service. Probationary teachers were evaluated at least once a year in more than 98% of the school districts responding to the ERS survey, but tenured teachers were evaluated less than once a year in the majority of the school districts. Most states stipulate the frequency of evaluations in state statutes. The states of Texas, California, and Ohio, for example, require annual evaluations of both probationary and tenured teachers. Distinctions based on employment status (probationary or tenured) need to be addressed in the procedures. In addition, individual teachers should be notified that they will be evaluated in a given year for the sake of clarity.

The principal is the supervisor who typically conducts evaluations of staff within his or her building (in more than 93% of the schools responding to the ERS survey), with observations by assistant principals and department heads in some cases. Some states require that only certified administrators conduct evaluations, but unless a state statute stipulates otherwise, anyone in the school system with the skills and training to make assessments could fulfill that role. Ideally, evaluators should be trained in the chosen assessment techniques and should be familiar with the overall evaluation system (Beckham, 1985a; Frels et al., 1984). Stronge and Tucker (1995) found that 36% of the states responding to their survey provided state-sponsored formal training in teacher evaluation for administrators. In addition, an increasing number of local school systems are addressing the issue of better training for supervisors (Petrie & Black, 1983). The ERS (1988) found that 84.8% of the local school districts surveyed provided training to evaluators prior to their assessing teacher performance. Ten years earlier, only 61.4% of the school systems had provided similar training. The increased effort to provide more training, coming from the state level in some cases, was attributed to concerns with the credibility and effectiveness of teacher evaluation (ERS, 1988).

In addition to identifying the participants in the evaluation process, evaluation procedures should provide an approximate timeline for the evaluation process, forms to be used, the steps to be taken (minimum number of observations), and how they will be accomplished (Frels et al., 1984; Petrie & Black, 1983). Although the courts have not required school systems to adhere to every aspect of their written evaluation procedures in cases of dismissal, teachers do reasonably expect that the written procedures will guide the process for the most part. Therefore, it is important that the evaluation procedures be realistic and somewhat flexible in nature. For example, a range of dates versus a rigid timeline provides adequate notice for teachers and yet allows for unexpected events and conflicts.

The majority of schools still rely on observation as the primary data collection method for teacher evaluation, although alternative methods are being tried in some school systems (ERS, 1988). As noted by Frels et al. (1984), it would be difficult to "justify and defend an evaluation of a teacher's performance that did not include some classroom observation" (p. 10). But the use of additional data sources increases both the validity and the legal credibility of a teacher evaluation, and it is especially important in cases of unsatisfactory evaluation. In *Rosso v. Board of School Directors* (1977), the court commended the school system's "model" evaluation procedures in a dismissal case wherein the principal, the superintendent, and three other administrators observed and rated the teacher's performance with similar results. The court found that multiple perspectives lessened the influence of personal bias and prejudice. The court also noted that the superintendent's method of recording what was going on in the classroom at 5-minute intervals gave the "best picture of the learning atmosphere in a classroom that [the court had] seen to date in the anecdotal record" (p. 1330).

Evaluation Summary

The evaluation summary should provide a comprehensive history of the teacher's performance relative to all of the major evaluation criteria. It should note improvement and lack of improvement in specified areas of performance. Examples of specific events and factual information should be used to support conclusions, especially in cases of outstanding or unsatisfactory performance. The focus of the summary should be on patterns of behavior, not isolated incidents, and the impact of that behavior on students or others (Stronge & Helm, 1991). Any personnel decisions made, either positive or negative, should be substantiated by the summarized record of performance contained in the report (Frels et al., 1984).

What Are the Implementation Aspects of an Evaluation System?

Most school districts have broad latitude in how they shape their evaluation procedures, but it is imperative in the implementation that all teachers be informed of the procedures and that all teachers be treated the same. Less than uniform treatment of all teachers could suggest arbitrary treatment, especially when found in conjunction with evidence of discrimination; this variable treatment would be actionable by a teacher (Beckham, 1985a, p. 8). As suggested by Frels et al. (1984), teachers should be treated as the principal would want to be treated if he or she was in the same position, and a climate should be established in which the personal dignity of teachers is preserved.

Just as a curriculum is no better than the teachers who deliver it, an evaluation system is no better than the administrators/evaluators who implement it. Implementation requires ongoing decision making and wisdom in striking the proper balance between encouraging professional growth and requiring accountability for student learning. Effective implementation requires consideration of the following elements:

- Documentation
- Notice of performance
- Recommendations for improvement
- Time frame for remediation
- Grounds for dismissal

Documentation

The primary purpose of the documentation process in most evaluation systems is to provide teachers with useful feedback on their professional responsibilities within the school setting. However, in cases of unsatisfactory performance, the documentation process serves as an important record of the school's actions in a personnel matter. Although documentation is not required on every aspect of an unsatisfactory teacher's performance, it is important that a sufficient sampling of the problems be documented at the time they occur so that it is not necessary to rely only on recall at a later date (Frels et al., 1984). The cumulative weight of relevant and credible evidence is needed to justify a negative personnel decision.

A relevant and credible documentation system should include (a) routine memoranda to the file (minor events of note), (b) specific incidents memoranda (noteworthy events or behavior), (c) visitation memoranda, and (d) summary memoranda (Frels & Horton, 1994). Such records provide principals with factual information on which to base final evaluations, provide teachers with adequate notice regarding potential problem areas, and provide concrete evidence in the event of a termination. To achieve these goals, the principal must engage in ongoing communication with the teacher through conferences and follow-up copies of file memoranda. In fact, Frels and Horton (1994) state that "the cardinal rule of documentation is to conference first and write second" (p. 8), because if an event is important enough to document, it is important enough to warrant seeking the teacher's perception of the situation. A teacher should also be given an opportunity to respond in writing with a memorandum to the file if he or she disagrees with the principal's comments.

Good communication skills are central and fundamental to the evaluation process at all stages (McGrath, 1993; Stronge & Helm, 1991). Frels and Horton (1994) recommend that communication be "direct, to the point, and educational jargon should be avoided" (p. 9). The more prompt and specific the feedback to the teacher, the greater the opportunity for improvement and avoidance of a termination. Specific feedback is facilitated by the collection of factual data reflecting teacher behaviors and consequences versus judgments and conclusions that should be reserved for the final summative evaluation. According to the court in *Gwathmey v. Atkinson* (1976), a termination "decision must be based upon fact and supported by reasoned analysis" (p. 1117).

Effective documentation, ideally, "substantiates the fundamental fairness of the process, establishes a reasonable predicate for the employment decision, and elaborates the procedural integrity of the process" (Beckham, 1985a, p. 1). It can be used to support the integrity of a termination decision even if there are possible charges of discrimination or free speech violations. If there is adequate documentation of deficiencies despite free speech activity, such as involvement in union activities, remediation can continue and termination remains an option if necessary. When evaluations are used as evidence in litigation, the judicial standard is that the documentation be "reliable, probative [i.e., must actually prove the facts at issue], and substantial" (Petrie & Black, 1983, p. 1050).

Notice of Performance

Whatever the source or sources of information used to judge teacher performance, a written record of the information should be shared with the teacher within a short period of time so that he or she is informed of problems and given a chance to respond and improve if necessary (*Needleman v. Bohlen*, 1979). In the case of classroom or other similar observations, specific facts, not conclusions, should be noted (Frels et al., 1984). Factual reporting based on observation and other sources allows for more specific and focused commendations for superior performance as well as suggestions for improvement in cases of unsatisfactory performance. Short, informal observations can be used to supplement the more traditional, full-class-period observation. Other sources of information might include parent surveys, student surveys, peer evaluations, and portfolios (Bridges, 1992; Peterson, 1995; Stronge & Helm, 1991).

When teachers are informed of both the positive and negative findings of the evaluation process, areas of weakness especially must be clearly identified and their relative importance to the overall evaluation must be explained. Teachers should be advised of the fact that failure to improve may lead to termination. In West Virginia, the court of appeals reinstated a probationary teacher due to a board of education's failure to inform the teacher how he had performed in his job (*Wilt v. Flanigan*, 1982). However, in *Rogers v. Department of*

Defense (1987), the issuance of a warning letter that stated the critical nature of four deficiencies that needed to be corrected within 2 months to achieve a satisfactory performance level was viewed as proper notice procedure by the court. In the event of unsatisfactory improvement, the specific consequences must be delineated.

Teachers must also be given an opportunity to respond to the grounds for dismissal by presenting evidence on their own behalf prior to discharge (*Cantrell v. Vickers*, 1980; *Needleman v. Bohlen*, 1979). A major case addressing this issue was brought on behalf of terminated school employees in Ohio, challenging the proper accordance of their due process rights. The U.S. Supreme Court, on *certiorari*, held that terminated employees were entitled to an opportunity to respond to charges against them (*Cleveland Board of Education v. Loudermill*, 1985).

Recommendations for Improvement

Once unsatisfactory performance has been documented, weaknesses need to be identified, a growth plan developed with specific directives for improvement, and assistance in meeting the directives provided (Frels & Horton, 1994; Frels et al., 1984). This is usually done with both oral and written communications to ensure that the teacher understands precisely what behaviors need to be improved, what behaviors are expected, and what the consequences will be if the teacher does not improve in the specified time period. These steps should provide a teacher with reasonable opportunity to improve and substantiate fair treatment by the principal. Any deficiencies that have not been remedied from previous years should be included in the recommendations, or it can be assumed that they are no longer a concern.

In *Iverson v. Wall Board of Education* (1994), a tenured teacher of 10 years was rated unsatisfactory by her new principal, and she brought charges based on personal bias and a violation of policy regarding the lack of a plan of assistance. The Supreme Court of North Dakota found that the principal had not acted arbitrarily, capriciously, or unreasonably, but he had failed to provide the teacher with adequate opportunity to improve due to the absence of a plan of assistance. Some states specifically require remediation assistance and others do not, but the courts typically expect it as a component of due process.

Time Frame for Remediation

As Adams (1988/1989) has noted, "Very little research on the effectiveness of remediation programs for teachers has been reported to date" (p. 24). Frase (1992) suggests that "only 10% of the teachers found to be incompetent ever achieve competency" (p. 70). Despite this dismal projection, it is important for schools to provide an opportunity for and assistance in remediation to demonstrate a commitment to fairness. Redfern (1983) and later Frase (1992) have recommended similar four-phase processes that are thorough and fair, providing the unsatisfactory teacher ample opportunity to seek assistance and to improve, if possible. The process described by both authors consists of the following steps: (a) early diagnosis, (b) performance improvement plan, (c) notification of corrective action, and (d) implementation of termination. The entire process takes 28 to 34 weeks of focused attention and documentation. Although this is more prolonged than what is probably required, it ensures full due process safeguards.

State statutes typically require that a reasonable time period be allowed for the remedy of a deficiency. "Reasonableness" will depend on the specific facts of the case, such as the nature of the deficiency and its impact on children. The Federal Circuit Court upheld a termination of a tenured teacher of 8 years that provided only 2 months for remediation in *Rogers v. Department of Defense* (1987), but the Supreme Court of Minnesota, in *Ganyo v. Independent School District* (1981), disallowed 8 weeks as a reasonable time for a teacher of 17 years to remedy teaching practices that had been labeled deficient for the first time. In *Rogers*, there had been ongoing difficulties and explicit identification of the expected performance criteria, whereas in *Ganyo*, there was no history of problems. These two cases illustrate how different circumstances can render highly discrepant court rulings. Whereas 2 months was deemed an acceptable time frame for remediation in *Rogers*, a year, as suggested by Frase (1992) and Redfern (1983), is probably a more realistic period.

BEST COPY AVAILABLE

Grounds for Dismissal

Although dismissals are necessary in a very small minority of cases, a majority of school districts are involved in them on a regular basis. In the 1988 ERS survey, 71.9% of the school districts reported the resignation or termination of a probationary teacher for an unsatisfactory evaluation during the previous 2 years, and 44.2% of the school systems reported similar experiences with a tenured teacher. Once a decision for termination has been made, the teacher should be advised of this decision and of his or her rights to appeal, informed of the possible consequences of a resignation, and given an opportunity to resign (Frels & Horton, 1994). Because teachers are often afforded the choice to resign, unsatisfactory evaluations frequently result in resignations rather than terminations.

Grounds for dismissal are set forth in most state statutes and typically include incompetency or inefficiency, insubordination, immorality, and failure to improve deficiencies as requested. Specific examples of the grounds for dismissal from the three states of California, Ohio, and Texas are given in Table 3.2. Incompetency or inefficiency is often cited as a reason for termination or demotion, and the terms typically are used interchangeably to refer to the "lack of some requisite ability" (Tigges, 1965, p. 1094). Incompetency or inefficiency is assumed to entail a lack of both the required knowledge to teach a given subject and the instructional skills to impart that knowledge. The terms may also be used in reference to other areas of responsibility, such as classroom management, student assessment, and duties outside the classroom. Incompetency or inefficiency "ordinarily manifests itself in a pattern of behavior, rather than in a single incident" (Tigges, 1965, p. 1095). Thus, it is essential to have an evaluation system that can document patterns of behavior. Failure by the teacher to remedy deficiencies once identified constitutes permissible grounds for dismissal (*Community Unit School District No. 60 v. Maclin*, 1982).

In summary, when an evaluation system is implemented, the evaluation procedures should be communicated to teachers before the process is begun and should be applied uniformly and systematically (Beckham, 1985b). The use of multiple sources of information based on different data collection strategies or different evaluators is important for reducing bias and corroborating findings (Bridges,

TABLE 3.2 Examples of Statutory Grounds for Dismissal

California (*California Education Codes Annotated*, 1993b/1995, § 44932)

1. immoral or unprofessional conduct
2. commission, aiding, or advocating the commission of acts of criminal syndicalism
3. dishonesty
4. unsatisfactory performance
5. evident unfitness for service
6. physical or mental condition unfitting him or her to instruct or associate with children
7. persistent violation of or refusal to obey the school laws of the state or reasonable regulations
8. conviction of a felony or of any crime involving moral turpitude
9. violation of Section 51530 of this code
10. knowing membership by the employee in the Communist Party
11. alcoholism or other drug abuse which makes the employee unfit to instruct or associate with children

Ohio (*Ohio Revised Code Annotated*, 1995, § 3319.16)

1. gross inefficiency or immorality
2. willful and persistent violations of reasonable regulations of the board of education
3. other good and just cause

Texas (*Texas Education Code Annotated*, 1991/1995, §§ 13.109, 13.110)

1. immorality
2. conviction of any felony or other crime involving moral turpitude
3. drunkenness
4. physical or mental incapacity preventing performance of the contract of employment
5. repeated and continuing neglect of duties
6. incompetency in performance of duties
7. failure to comply with reasonable requirements prescribed by the school district for achieving professional improvement and growth
8. willful failure to pay debts
9. habitual use of addictive drugs or hallucinogens
10. excessive use of alcoholic beverages
11. necessary reduction of personnel
12. for good cause
13. failure to perform satisfactorily on examination

1992; Peterson, 1995). The preliminary specification of job-related criteria, fair notice of the implementation procedures and evaluation findings, opportunities to discuss and remediate identified deficiencies, and a thorough documentation system constitute a fair and reasonable evaluation system. An evaluation system that embraces the principles described above is one that is likely to be viewed as fair to the teacher and as according due process safeguards while also supporting a termination decision if necessary.

What Are the General Principles for Teacher Evaluation?

In essence, the due process and equal protection clauses of the Fourteenth Amendment protect individuals from discriminatory practices by the state. Statutory law further defines the rights of individuals under state law. Compliance with the law and fundamental fairness can be achieved through the adoption of professionally relevant standards of performance and the uniform application of evaluation procedures (Beckham, 1985b; Joint Committee on Standards for Educational Evaluation, 1988).

Most legal challenges have focused on the procedural aspects of the personnel evaluation process versus actual findings (Adams, 1988/1989; Cain, 1987; Stronge & Helm, 1991). The courts typically have been reluctant to overrule the judgment of a supervisor unless there has been a lack of sufficient documentation to support the conclusions reached, the decisions are deemed "arbitrary or capricious," or there has been evidence of retribution for actions unrelated to job performance, such as the exercise of free speech regarding public issues. As noted in *Parker v. Board of School Commissioners of Indianapolis* (1984), the court's task "is not to assess the 'wisdom' of employment decisions, but rather to determine whether an employer's selection procedures comport with the law" (p. 529).

The legal context for teacher evaluation provided by the foregoing review of key aspects of the U.S. Constitution, state statutes, and case law collectively suggests the principles summarized in Table 3.3 to guarantee procedural due process in teacher evaluation. These principles not only meet the general legal requirements of a fundamentally fair evaluation process, they also provide the framework for one that is ethical and supportive of professional growth.

TABLE 3.3 Legal Principles for Teacher Evaluation

1. *Written policy:* The criteria for evaluation and the exact procedures to be followed should be formalized in a written policy.
2. *Legal compatibility:* Any written policy should satisfy state statutory mandates and collective bargaining agreements.
3. *Prior notification:* All employees must be fully informed of the criteria for evaluation and the exact procedures to be followed before the evaluation process begins.
4. *Trained evaluators:* Evaluators should be knowledgeable about job requirements and trained to collect data in the manner stipulated by policy.
5. *Job-related criteria:* Evaluation criteria should be as objective as possible, reflect actual job requirements and school system goals, and be validated in some manner.
6. *Equitable treatment:* Evaluation procedures should treat all employees in an equitable manner as prescribed by state and local requirements.
7. *Pattern of performance:* The evaluation procedures should assess a broad sampling of each employee's job responsibilities over a sufficient period of time to identify a pattern or patterns of behavior.
8. *Documentation:* A record of all written communications regarding the performance of a teacher should be maintained.
9. *Opportunity to improve:* If job performance is unsatisfactory, there must be adequate notice of concerns, clear directions for improvement, reasonable time for improvement, and assistance in improvement. (Statutory law clearly defines this aspect of due process in some states.)
10. *Substantiation for decisions:* Evaluation summaries must be consistent with and provide sufficient documentation to substantiate personnel decisions.

Conclusion

Although many administrators are intimidated by the legal requirements inherent in personnel evaluation, the basic principles set forth in law provide a fundamentally fair and rational system for designing a sound evaluation system and making personnel decisions based on the proper implementation of the system. The courts have supported school districts' personnel decisions when teachers have been accorded fundamental due process rights. The protection of these rights requires time and commitment of resources to the evaluation process, but the provision of quality education demands nothing less. The professional development of the vast majority of teachers and the dismissal of the few incompetent teachers are well worth this heightened level of commitment and effort to provide meaningful evaluation. A legally defensible evaluation system is one that provides a useful process for the professional development of most teachers and is ethically sound and fair to all parties.

Note

1. Based on the rational basis standard or test, "an appellate court will not second-guess the legislature as to the wisdom or rationality of a particular statute if there is a rational basis for its enactment, and if the challenged law bears a reasonable relationship to the attainment of some legitimate governmental objective" (Black, 1991, p. 872).

References

Adams, F. E. (1989). Legal aspects of teacher dismissal for incompetence in the public schools (Doctoral dissertation, University of Alabama at Birmingham, 1988). *Dissertation Abstracts International, 50,* 311.

Alexander, K., & Alexander, D. M. (1985). *American school law* (2nd ed.). St. Paul, MN: West.

Alkin, M. C. (1992). *Encyclopedia of educational research* (6th ed.). New York: Macmillan.

Beckham, J. C. (1985a). *Legal aspects of employee assessment and selection in public schools.* Topeka, KS: National Organization on Legal Problems of Education.

Beckham, J. C. (1985b). Legally sound criteria, processes and procedures for the evaluation of public school professional employees. *Journal of Law and Education, 14,* 529-551.

Black, H. C. (1991). *Black's law dictionary.* St. Paul, MN: West.

Board of Regents v. Roth, 408 U.S. 564 (1972).

Bridges, E. M. (1992). *The incompetent teacher: Managerial responses.* Washington, DC: Falmer.

Cain, B. L. (1987). Nonformal teacher discipline (Doctoral dissertation, University of Houston, 1987). *Dissertation Abstracts International, 48,* 1066-1067.

California Education Codes Annotated. (1993a). §§ 33039, 44662-44664. St. Paul, MN: West.

California Education Codes Annotated. (1993b). § 44932 (West Supp. 1995). St. Paul, MN: West.

Cantrell v. Vickers, 495 F. Supp. 195 (N.D. Miss. 1980).

Cleveland Board of Education v. Loudermill, 470 U.S. 532 (1985).

Community Unit School District No. 60 v. Maclin, 106 Ill. App.3d 156, 435 N.E.2d 845 (Ct. App. 1982).

Conley v. Board of Education, 143 Conn. 488, 123 A.2d 747 (1956).

Corcoran, T., & Goertz, M. (1995). Instructional capacity and high performance schools. *Educational Researcher, 24*, 27-31.

Crone, L. J., & Teddlie, C. (1995). Further examination of teacher behavior in differentially effective schools: Selection and socialization process. *Journal of Classroom Interaction, 30*(1), 1-9.

Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 17-34). Newbury Park, CA: Sage.

Delon, F. G. (1982). *Legal issues in the dismissal of teachers for personal conduct*. Topeka, KS: National Organization on Legal Problems of Education.

Educational Research Service (ERS). (1988). *Teacher evaluation: Practices and procedures*. Arlington, VA: Author.

Eiche v. Louisiana Board of Elementary and Secondary Education, 582 So.2d 186 (La. 1991).

Eshom v. Board of Education of School District No. 54, 364 N.W.2d 7 (Neb. 1985).

Frase, L. E. (1992). *Maximizing people power in schools*. Newbury Park, CA: Corwin.

Frels, K., Cooper, T. T., & Reagan, B. R. (1984). *Practical aspects of teacher evaluation*. Topeka, KS: National Organization on Legal Problems of Education.

Frels, K., & Horton, J. L. (1994). *A documentation system for teacher improvement or termination* (4th ed.). Topeka, KS: National Organization on Legal Problems of Education.

Fuhr, D. L. (1993). Managing mediocrity in the classroom. *School Administrator, 50*(4), 26-29.

Fuhrman, S. H. (1994). Legislatures and education policy. In R. F. Elmore & S. H. Fuhrman (Eds.), *The governance of curriculum* (1994 ASCD yearbook) (pp. 30-55). Alexandria, VA: ASCD.

Ganyo v. Independent School District No. 832, 311 N.W.2d 497 (Mn. S.Ct. 1981).

Gwathmey v. Atkinson, 447 F. Supp. 113 (E.D. Va. 1976).

Harris, W. U. (1981). Teacher command of subject matter. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 58-72). Beverly Hills, CA: Sage.

Iverson v. Wall Board of Education, 522 N.W.2d 1988 (S.D. 1994).

Iwanicki, E. F. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 158-171). Newbury Park, CA: Sage.

Johnson, S. M. (1984). *Teacher unions in schools*. Philadelphia: Temple University Press.

Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to assess systems of evaluating educators*. Newbury Park, CA: Sage.

Jones v. Mt. Healthy Board of Education, No. A-7305247 (C.P. Hamilton County, Ohio, April 8, 1983).

Laing, S. O. (1986). The principal and evaluation. *NASSP Bulletin, 70*(493), 91-93.

Long v. School District of University City, 777 S.W.2d 944 (Mo. App. 1989).

McCarthy, M. M., & Cambron-McCabe, N. H. (1992). *Public school law: Teachers' and students' rights* (3rd ed.). Boston: Allyn & Bacon.

McGrath, M. J. (1993). When it's time to dismiss an incompetent teacher. *School Administrator, 50*(3), 30-33.

McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher evaluation: Improvement, accountability, and effective learning*. New York: Teachers College Press.

Millman, J. (1981). Introduction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 12-13). Beverly Hills, CA: Sage.

Montgomery v. Trinity Independent School District, 309 F.2d 1058 (5th Cir. 1987).

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Government Printing Office.

Needleman v. Bohlen, 602 F.2d 1 (1st Cir., 1979).

Nowak, J. E., & Rotunda, R. D. (1995). *Constitutional law*. St. Paul, MN: West.

Ohio Revised Code Annotated. (1995). §§ 3319.11-3319.16. Cleveland: Banks-Baldwin Law.

Parker v. Board of School Commissioners of Indianapolis, 729 F.2d 524 (7th Cir. 1984).

Perry v. Sindermann, 408 U.S. 593 (1972).

Peterson, K. D. (1995). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin.

Petrie, B. I., Jr., & Black, T. S. (1983). Teacher termination in Ohio for "gross inefficiency": The role of teacher evaluation. *Ohio State Law Journal, 44*, 1041-1050.

Redfern, G. B. (1983). Dismissing unsatisfactory teachers: A four phase process. *ERS Spectrum, 1*(2), 17-21.

Rogers v. Department of Defense, 814 F.2d 1549 (Fed. Cir. 1987).

Rosso v. Board of School Directors, 33 Pa. Cmwlth. 175, 380 A.2d 1328 (1977).

Rossow, L. F., & Parkinson, J. (1992). *The law of teacher evaluation*. Topeka, KS: National Organization on Legal Problems of Education.

Sargent v. Selah School District No. 119, 599 P.2d 25 (Wash. App. 1979).

Sclan, E. M. (1994, July). *Performance evaluation for experienced teachers: An overview of state policies*. Paper presented at the annual conference of the National Evaluation Institute, Gatlinburg, TN.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

Shapiro, B. C. (1995). National standards for teachers. *Streamlined Seminar, 13*(4). (ERIC Document Reproduction Service No. ED 381 502)

Stronge, J. H., & Helm, V. M. (1991). *Evaluating professional support personnel in education*. Newbury Park, CA: Sage.

Stronge, J. H., & Tucker, P. D. (1995). Performance evaluation of professional support personnel: A survey of the states. *Journal of Personnel Evaluation in Education, 9*, 123-137.

Texas Education Code Annotated. (1991). §§ 13.047, 13.103-13.115, 13.302-13.306, 21.201-21.205 (Vernon Supp. 1995). St. Paul, MN: West.

Tigges, J. H. (1965). What constitutes "incompetency" or "inefficiency" as a ground for dismissal or demotion of public school teacher. In *American Law Reports* (4 ALR 3d. 1090). Rochester, NY: Lawyers Co-operative.

United States v. South Carolina, 445 F. Supp. 1094 (D.S.C. 1977), *aff'd*, 434 U.S. 1026 (1978).

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249-294.

Wethersfield Board of Education v. Connecticut State Board of Labor Relations, 519 A.2d 41 (Conn. 1986).

Wilt v. Flanigan, 294 S.E.2d 189 (W.V. Ct. App. 1982).

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices.* Santa Monica, CA: RAND Corporation.

Yudof, M. G., Kirp, D. L., & Levin, B. (1992). *Educational policy and the law.* St. Paul, MN: West.

4

Applying the Personnel Evaluation Standards to Teacher Evaluation

JAMES R. SANDERS

Standards in education have played a major role in the evaluation of schools, educational programs, student performance, and school personnel over the past century. As educators, we have available to us accreditation standards, subject-matter standards, test standards, and evaluation standards. The purpose of this chapter is to take one set of standards, the Personnel Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (1988), and look at their applicability to teacher evaluation.

The chapter begins with background information about the Standards and a description of the process used in their development. This introduction is then followed by a description of the content of the Standards and its implications for teacher evaluation. I conclude the chapter with a final word about the importance of using the Standards as a foundation for developing teacher evaluation systems.

Background

The Joint Committee on Standards for Educational Evaluation was created in 1975 in response to a growing need to establish expectations for evaluations in education. It is a coalition of 16 professional associations that meet annually to review issues in educational evaluation.[1] When there has been a need for setting standards, the Joint Committee has requested funds from foundations, government, and other sources to develop the necessary standards. The Joint Committee has published two sets of standards for evaluations that are now widely accepted in education. The first version of the Personnel Evaluation Standards was published in 1988; the second was published in 1994.

Actually, the first standards published by the Joint Committee were standards for the evaluation of educational programs, projects, and materials (1981). This document was the first to describe generally accepted evaluation practices applied to educational programs. These standards were replaced in 1994 by the Program Evaluation Standards. Soon after the 1981 publication came out, the Joint Committee tackled the much more complex problems of teacher, and other educator, evaluation. This effort resulted in the Personnel Evaluation Standards (1988), which describe generally accepted personnel evaluation practices in education. Many school districts and one state (Hawaii) have adopted these standards as a foundation on which to build a defensible teacher evaluation system. A national research center, the Center for Research on Educational Accountability and Teacher Evaluation, has based its work on the Personnel Evaluation Standards. Even the U.S. Marine Corps has used these standards to review its own personnel evaluation practices.

The Personnel Evaluation Standards have widespread utility and address very important needs. There is a need to evaluate entry-level teachers before certification. There is a need for guidance in hiring decisions, promotion and tenure decisions, and the recognition and reward of meritorious contributions. There is a need to help faculty and administrators take stock of teacher strengths and weaknesses and develop professional development plans. And there is a need to develop fair, valid, and effective cases when it becomes necessary to terminate teachers who are harmful to students or otherwise ineffective. The need for quality teacher evaluation is important, pervasive, and multifaceted (Joint Committee, 1988).

At the time when the Personnel Evaluation Standards were published, there was widespread concern about the quality of teacher evaluation (Joint Committee, 1988, p. 6, App. A). One group of researchers had noted that teacher evaluations were often "subjective, unreliable, open to bias, closed to public scrutiny, and based on irrelevancies" (Soar, Medley, & Coker, 1983, p. 246). This conclusion was supported by research conducted by the RAND Corporation that concluded that most teacher evaluation systems were illogical, simplistic, unfair, counterproductive, or unproductive (Darling-Hammond, Wise, & Pease, 1983). The Joint Committee concluded that a major problem was that there was no widespread agreement about the characteristics of good personnel evaluation practices, and it set out to fill that void.

How Were the Personnel Evaluation Standards Developed?

The Joint Committee has developed a set of operating procedures that guide its work in setting standards (Sanders, 1994). A fundamental piece of the whole standards-setting process is the way in which the Joint Committee has defined the concept of a standard. According to the Joint Committee, a standard is a principle commonly agreed to that is used to assess the quality of an evaluation. The key phrase is "a principle commonly agreed to."

The standards-setting process used by the Joint Committee is open, public, and participatory. Its goal is to describe evaluation practices that all will agree are appropriate and acceptable. To achieve this goal, the Joint Committee takes the following steps in setting standards:

1. *Identification of issues:* Educators and others are invited to identify problems that need to be addressed. A review of the literature is used to supplement reported issues.
2. *Development of a first draft:* A panel of approximately 50 writers who are experienced in the topics to be covered by the standards are commissioned to develop initial versions of the standards. Each topic has at least two different people writing to a common outline. These drafts are reviewed, critiqued,

and rewritten by members of the Joint Committee into a first draft of the standards.

3. *National and international reviews:* The first draft is critiqued by at least 60 reviewers identified by the member associations. The draft is also sent for review to anyone else who requests it. The criteria used for review include (a) the need for the document, (b) responsiveness to concerns in the field, (c) courage, (d) validity of the content, (e) practicality, (f) legality, (g) clarity, and (h) depth of treatment. Based on the reviews, a second draft of the standards is created by the Joint Committee.
4. *Field tests:* Trial use of the second draft is arranged through the member associations in about 40 different settings and uses. Following those pilot tests, reports are provided to the Joint Committee to guide its development of the third draft of the standards.
5. *National public hearings:* Hearings are organized to provide an open forum for anyone who wishes to comment on the standards. Those who cannot provide testimony in person are encouraged to submit written comments. The testimony is reported to the Joint Committee for use in developing a fourth draft.
6. *ANSI review:* The fourth draft is submitted to the American National Standards Institute for announcement in *Standards Action*, a bulletin that goes to all who are interested in the setting of voluntary standards. Any reactions require response by the Joint Committee.
7. *Finalization of the standards:* Once the final draft of the standards is ready, the Joint Committee reviews, revises, and votes on the final version and submits the document for publication.

It should be obvious from this process that anyone who has anything to say about the substance or format of the standards has ample opportunity over several years of development to contribute. A validation panel of educators oversees the process to ensure that the procedures have been followed faithfully. The resulting product is a compilation of best practices and warnings against common errors, as found in the literature and in use in education at the time. These standards are subject to review every 5 years to ensure that they are kept up-to-date. There is no better source to go to for use in reviewing current practices in schools.

What Are the Implications of the Personnel Evaluation Standards for Teacher Evaluation?

The Personnel Evaluation Standards are organized into four categories that reflect the attributes of sound and fair teacher evaluation practice: propriety, utility, feasibility, and accuracy. There are several standards within each of these categories that indicate how the category can be developed. The standards are the means for developing sound personnel evaluation practices.

Propriety

The propriety standards establish an expectation that teacher evaluations will be conducted legally, ethically, and with due regard for the welfare of those being evaluated and those they serve. If there are questions of propriety in any evaluation, the Standards call for resolving them before going any further.

- The *service orientation* standard states that teacher evaluations should promote sound education principles, fulfillment of school purposes, and effective performance of job responsibilities so that the educational needs of students, community, and society are met. Teacher evaluations should protect student and community rights to equal educational opportunity and competent instruction. Educational systems exist to meet the needs of students and the community. All elements of those systems, including teacher evaluation, should be directed toward achieving that purpose. Standards of teaching need to be set and enforced consistently for all teachers.
- The *formal evaluation guidelines* standard calls for written policies, purposes, procedures, and criteria to enable teacher evaluations to be conducted in a consistent, equitable, and legal manner. Written guidelines protect both the school district's and the teacher's rights. Written guidelines should be consistent with existing laws, contracts, and school district policies.

- The *conflict of interest* standard states that any teacher evaluation process should guard against evaluations that are inappropriately influenced by personal self-interest rather than school district or professionally shared goals. The teacher evaluator should guard against inappropriate personal biases influencing judgment or decisions about a teacher.
- The *access to personnel evaluation reports* standard calls for limiting access to teacher evaluations to the teacher and to those who are authorized to make decisions and take actions based on the evaluations. Teacher evaluations are confidential. Access to them should be limited and controlled.
- The *interactions with evaluatees* standard states that teachers being evaluated should be treated in a professional, considerate, and courteous manner. Teacher morale and performance can be affected by a failure to address this standard.

Utility

The utility standards call for teacher evaluations that are informative, timely, and influential. If personnel evaluations are not useful, one must question why they are being done.

- The *constructive orientation* standard calls for teacher evaluations that help school districts develop their human resources and assist teachers to excel in their responsibilities. Teacher evaluations should promote the success of students, teachers, and school districts. Examples of constructive teacher evaluations include those that reinforce good practice, provide direction for improvement, promote professionalism, foster collegiality and harmony, and assist in terminating incompetent teachers.
- The *defined uses* standard states that the users and intended uses of teacher evaluations should be clearly identified at the start. Adherence to this standard should lead teacher evaluations to answer the questions of intended users. It also states that teacher evaluations should not be used for purposes other than those intended unless there is concurrence by all involved parties.

- The *evaluator credibility* standard establishes an expectation that teacher evaluations should be managed and conducted by qualified, trained, authorized, supported, and monitored people.
- The *functional reporting* standard calls for clear, timely, accurate, and germane reporting of teacher evaluations. These results should have practical value for those identified as intended users.
- The *follow-up and impact* standard calls for assistance to intended users in using the results of teacher evaluations to take appropriate actions. Merely reporting teacher evaluation findings is insufficient.

Feasibility

The feasibility standards are concerned with promoting teacher evaluations that are easy to implement, efficient, adequately supported, and viable from the perspectives of stakeholders. Any personnel evaluation system should be realistic and workable. We do what we can and then strive continuously to improve practices over time.

- The *practical procedures* standard calls for minimal disruption and cost in conducting teacher evaluations. Practicality should not override accuracy, but procedures should be appropriate for the information required and the resources available.
- The *political viability* standard calls for teachers, evaluators, and other intended users to collaborate in establishing goals, policies, and procedures for teacher evaluation that are acceptable to all. Other intended users include school boards and potential external users such as parents, other citizens, and state certification and accreditation agencies.
- The *fiscal viability* standard establishes an expectation that adequate time and resources should be provided for teacher evaluations. Resources include trained evaluators, support staff, administrative staff, teacher time, help of outside experts, space, facilities, materials, and equipment.

Accuracy

The accuracy standards are concerned with the fidelity of teacher evaluations. Conclusions should be justified by reported evidence. We want to make sure that any teacher evaluation that is used to make personnel decisions is open to scrutiny and will pass close review. Attention to detail and precision, when personnel matters are being considered, is critical to the building of a sound personnel evaluation process.

- The *defined role* standard states that the role and responsibilities (or needed qualifications in the case of hiring decisions) of the teacher should be clearly defined for any teacher evaluation. All parties involved in a teacher evaluation should have the same understanding of the position requirements before proceeding with the evaluation. Performance expectations and position qualifications should be keyed to these requirements. Adherence to this standard creates a foundation for subsequent steps in the teacher evaluation process.
- The *work environment* standard recognizes that the context in which a teacher works can affect performance and, therefore, must be taken into account in teacher evaluations. Context may include school organizational structure, school district goals or objectives, curriculum mandates, leadership or supervisory practices, financial resources, student and family background, time, space, availability of instructional materials and equipment, support services, class size, and any number of calamities. Contextual influences should be referenced in a teacher evaluation to be fair, but should not be used to rationalize poor performance.
- The *documentation of procedures* standard calls for documenting the teacher evaluation process so that it can be checked if questions arise and improved if weaknesses are found. In particular, there should be documentation of steps, forms, appeal procedures, reporting, follow-ups, and due process.
- The *valid measurement* standard is the heart of the accuracy standards. Inferences drawn from measurements and inter-

pretations of evidence should be correct and trustworthy. Teacher evaluations should be assessed for their validity before actions are taken. *Validity* refers to whether evaluators are getting the information they think they are getting.

- The *reliable measurement* standard refers to the extent to which inconsistencies appear in the information collected about a single teacher characteristic. The validity of findings is affected by the reliability of measurements. *Reliability* refers to the precision with which information is reported and whether consistent information is being received.
- The *systematic data control* standard establishes an expectation that information collected for teacher evaluations will be kept secure and will be carefully processed so that errors are not introduced during the handling of data. Those who are charged with processing evaluations of teachers should be trained in the tasks they are expected to perform. Teachers should have an opportunity to review their records and to question perceived inaccuracies.
- The *bias control* standard calls for fair and impartial evaluation of teachers. An unbiased evaluation is one based solely on criteria and information relevant to the teacher's qualifications or performance.
- The *monitoring evaluation systems* standard tells school districts to review their teacher evaluation systems periodically and carefully so that they can be improved. A school district committee set up to review the teacher evaluation process every 1 to 2 years is essential. Outside reviewers are also highly recommended.

These standards provide a set of principles that should guide the design, development, and implementation of all teacher evaluation systems. These principles have been accepted by members of the Joint Committee, who represent most elementary and secondary educators in the United States. To summarize implications of these standards for teacher evaluation, a checklist of key questions is presented in Table 4.1. These questions may be used to review policies, procedures, content, and uses of teacher evaluations.

TABLE 4.1 Key Questions About Teacher Evaluations

Propriety

1. Will equal educational opportunity and competent instruction be served by the evaluation?
2. Are there clearly written guidelines for teacher evaluation?
3. Can personal biases influence the results of teacher evaluation?
4. Is access to teacher records limited and controlled?
5. Are teachers treated professionally and courteously?

Utility

1. Will the evaluation help the school district and the teacher excel in assigned responsibilities?
2. Are the users and intended uses of the evaluation clearly identified?
3. Are those who manage and conduct the teacher evaluation qualified?
4. Are the results of the evaluation of practical value and timely?
5. Is there assistance in using the results of the evaluation?

Feasibility

1. Are the procedures easy to implement, with minimal disruption and cost?
2. Were stakeholders in the teacher evaluation involved in the development of goals, policies, and procedures of teacher evaluation? Are the goals, policies, and procedures acceptable to all?
3. Are adequate time and resources provided for teacher evaluation?

Accuracy

1. Are the teacher's role, qualifications, and responsibilities clearly defined and understood?
2. Is there anything about the work environment that is affecting the teacher's performance and should be taken into account?
3. Have the evaluation steps, results, and actions been documented?
4. Are inferences made during the evaluation correct and trustworthy?
5. Are there inconsistencies on any measures?
6. Did the handling of information or records result in error being introduced? Are records secure?
7. Was the evaluation fair and impartial?
8. Has the system for teacher evaluation been reviewed and improved within the past year?

SOURCE: These questions are based on the Personnel Evaluation Standards published by the Joint Committee on Standards for Educational Evaluation (1988).

Suggestions for Using the Standards

Any school district that is looking at improving its teacher evaluation system will find the Personnel Evaluation Standards useful and indispensable. The Standards can be used in a variety of ways, but a suggested process for using them with a working group follows.

A task force or working group composed of teachers, teacher union representation, principals, central office staff, school board representation, older students, parents, and taxpayers may be established to represent the major stakeholders. Each member of the working group should be given a copy of the Standards and told to skim through them to see how they are organized.

At the first meeting of the working group, the group members should introduce themselves, and then the current system of teacher evaluation should be described. Supporting documents—such as forms, checklists, and sample reports used in the process—should be distributed. This presentation should then be followed by a brainstorming discussion listing the strengths (things to keep) and weaknesses (things to change) of the current system.

Members of the working group should then be assigned to solicit strengths, weaknesses, and issues from their respective groups (other teachers, principals, central office administrators, school board members, parents, taxpayers, students) and report back on these at the next meeting.

At the second meeting, a comprehensive listing of strengths, weaknesses, and issues should be compiled and typed up for future use. It is important that this compilation reflect both positive and negative practices.

At the third meeting, the working group should begin going through the standards one by one. After silently reading each standard and the guidelines and common errors associated with it, the group should discuss the strengths, weaknesses, and ideas for change that are suggested by the standard.

The working group may be able to discuss two standards at each meeting, requiring 11 meetings to complete the process. Once all standards have been discussed, the working group leader should prepare a summary of the changes desired in the teacher evaluation system. These changes can be discussed and voted on at the next meeting.

This planning period could take a year to complete. Once a plan is prepared, a second year could be used to pilot test and evaluate the system changes. Perhaps one school, or a few teachers and a principal, could volunteer to try out the changes and report back to the working group on the findings.[2] Based on this field test, final changes can be put into place before final adoption of the teacher evaluation system. This process then should be reviewed at least every 2 years, using the Standards as a guide.

A Final Word

The Personnel Evaluation Standards are a valuable tool for educational leaders who recognize that teacher evaluation is a necessary and important means for moving education toward excellence. Teacher evaluation gives direction to teachers and provides public accountability. The Standards are a compilation of best practices—the best that we know about how to conduct teacher evaluation. Use of the Standards in the manner suggested here will enable school districts to conduct teacher evaluation in a way that best benefits children and involves the important stakeholders in the educational process. If used as suggested, the Personnel Evaluation Standards also will enable us to conclude that the needs for better teacher evaluation that were identified in the early 1980s are being addressed.

Notes

1. The sponsoring organizations of the Joint Committee are the American Association of School Administrators, the American Educational Research Association, the American Evaluation Association, the American Federation of Teachers, the American Psychological Association, the Association for Supervision and Curriculum Development, the Canadian Evaluation Society, the Canadian Society for the Study of Education, the Council of Chief State School Officers, the Council on Recognition of Postsecondary Accreditation, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Council on Measurement in Education, the National Education Association, the National Legislative Program Evaluation Society, and the National School Boards Association.
2. Beginning teachers or administrators, or other high-consequence cases, should be used in the pilot test, because these are cases in which the Personnel Evaluation Standards can have the greatest impact.

References

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53,* 285-328.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects and materials.* New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to assess systems of evaluating educators.* Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards.* Thousand Oaks, CA: Sage.

Sanders, J. R. (1994). The process of developing national standards that meet ANSI guidelines. *Journal of Experimental Education, 63*(1), 5-12.

Soar, R. S., Medley, D. M., & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan, 65,* 239-246.

PART II

Assessing Teacher Performance

5

Classroom-Based Assessments of Teaching and Learning

CHAD D. ELLETT

Classroom observation, measurement, assessment, and evaluation are not new; they have seemingly always been integral parts of life in American schools. These processes have long been of interest to (a) educational policy makers concerned with questions about the means and ends of education; (b) researchers seeking to identify elements of effective teaching and learning; (c) programmatic efforts designed to improve schools and make them more effective; (d) teacher educators, supervisors, and others concerned with teacher mentoring and development; (e) school boards and school administrators charged with the responsibilities of teacher (employee) evaluation and educational accountability; (f) teacher organizations (i.e., teachers' unions) in their attempts to protect teacher rights; (g) state agencies concerned with teacher licensure; (h) those designing and implementing new models for teaching careers and compensation/incentive plans (e.g., career ladders, merit pay systems); and, most recently, (i) organizations concerned with the national professional certification and credentialing of teachers (e.g., the National Board of Professional Teaching Standards). Thus, it seems quite apparent that classroom-based assessments of teaching and learning have had, and continue to have, wide-ranging implications for the manner in which our culture views the educational enterprise.

In this chapter, I present a discussion of selected issues, methods, and procedures pertaining to classroom-based observations and assessments of teaching *and* learning. The focus is on alternative perspectives to those typically associated with traditional teacher evaluation practices. The chapter integrates a set of my own experiences in teacher evaluation and assessment over the past 20 years. It is grounded in a strong belief that the primary source of information useful for improving teaching and learning in classrooms stems from direct, systematic observation and assessment of teaching and learning using multiple, classroom-based methodologies. Thus, improving the quality of teaching and learning processes and general life in classrooms requires that principals, teachers, and perhaps students as well, attend to individual and collective assessments of actual classroom events that occur in an ongoing manner throughout the school year.

A Brief History of Classroom-Based Assessment of Teachers and Teaching

Through the years, as educational philosophies, beliefs, and values have changed, what has been valued in classrooms, and what has been observed, measured, assessed, and evaluated have changed as well. Early in the 20th century, for example, when ideas about teachers and good teaching were surrounded by issues and concerns pertaining to moral and practical education, considerable efforts were made to identify the traits that make teachers effective. According to Medley (1977), the Commonwealth Teacher Training Study (Charters & Waples, 1929) was the most outstanding and thorough study in its day that attempted to catalog the traits of effective teachers. Questions about teacher traits are not part of the pursuits of very many educational researchers today—and one would probably think it quite silly, not to mention illegal, to certify teachers based on scores attained on personality characteristics inventories created by the state. Official instruments and rating scales to be used in the evaluation of teachers began to appear around 1915, and they were used mainly by administrators (Medley, 1977). There was also concern expressed in the school administration and supervision literature that the primary role of school principals was to observe and assess teachers in the classroom and to work with them on the improvement of teaching. The following quote from a classic work on supervision written by Hubert Nutt in 1928 makes this point rather clear:

> The primary function of the school principal shall be to carry out an effective program of instructional supervision in his building. He shall devote not less than two thirds of the regular school day to personal visitation and study of the work of teachers in his building. . . . He shall select the number of teachers and the subjects which he will be able to work with intensively, and he shall visit each teacher not less than two class periods in each subject selected during each week for a period of not less than four weeks. At the end of this period of intensive supervision, he shall select either a new group of subjects with the same teachers, or a new group of teachers with the same subjects and continue as before. . . . He shall continue this plan throughout the school year. (p. 524)

The work of principals in today's classrooms remains important. However, the complexities of society in general and schools in particular almost vitiate the use of any form of assessment (in the spirit of classroom-based assessments as opposed to traditional evaluations) as described here. Subsequently, student ratings of teachers and teachers' self-ratings came on the scene (Good & Mulryan, 1990) as philosôphies of school administration began to move away from scientific management (Taylor, 1947) toward more humanistic perspectives.

During the 1940s and into the 1950s, teacher effectiveness became a concern of those looking into classrooms and evaluating teachers. Good and Mulryan (1990) cite the emergence of the *Ohio Teaching Record* in 1941 as making a major contribution to focusing classroom-based assessments on the study of teaching effectiveness (see Beecher, 1949). Subsequently, in the 1960s, 1970s, and 1980s, the era of teacher effectiveness research in American classrooms boomed. The availability of instruments for looking into classrooms using a variety of methodologies (e.g., time and event sampling, behavior checklists, category systems, sign systems, narrative records, rating scales) also rapidly proliferated, resulting in a variety of compilations of instruments and methodologies for conducting classroom observations for the purposes of analyzing teaching, promoting teacher development, and conducting teacher effectiveness research (e.g., Flanders, 1970; Medley, Coker, & Soar, 1984; Simon & Boyer, 1967; Stallings, 1977). These various instruments and procedures focused on a wide variety of macro- and microobservation and assessment variables (e.g., teacher warmth and enthusiasm versus the frequency of student questions) derived from various conceptions and models of good and/or effective teaching and generated rich resources on different models of teaching (e.g., Dunkin & Biddle, 1974; Gage, 1972) and for integrating classroom observation instruments and procedures with various teaching models (Stallings, 1977). As interests in measuring teacher behavior and teacher effectiveness soared, there were new calls for the development of "objective" and empirically based measures of teaching based on systematic classroom observation, and new insights were developing about establishing the psychometric properties of observation measures (Medley & Mitzel, 1963; Rosenshine & Furst, 1973).

One major result of the proliferation in classroom-based observation research in the 1970s and 1980s was the advent of the process-product research paradigm, which sought to identify teaching behaviors and classroom processes linked to student achievement. The results of process-product research, as well as cautions about overextrapolating research findings from this paradigm, exist in many sources (e.g., Brophy, 1986, 1988; Ellett, 1990; Gage & Needels, 1989; Medley, 1977; Porter & Brophy, 1988). One major effect of the emphasis given the process-product research paradigm during the 1970s and 1980s was a developing knowledge base about preferred criteria and procedures for developing classroom-based evaluations of teaching. There emerged a considerable shift in the thinking of local school districts about what was important to evaluate in teaching and how to conduct classroom observations to gather data to make judgments about effective teaching. Considerable attention was given to revising evaluation instruments and procedures for making summative evaluations of teachers to better align their content and focus with findings from the extant process-product literature. The Florida Performance Measurement System (FPMS) is probably the most prominent example of a system designed to rigidly reflect findings from the process-product literature and a philosophy of objectivism in teacher evaluation (e.g., see Medley et al., 1984).

At the same time, minimum competency testing of students and teachers and even school principals (Ellett, 1990) emerged in the 1970s, consistent with the national competency-based teacher education movement in the 1960s and early 1970s and the call for higher standards and educational accountability. Within this context, a variety of states began to question the limited utility and job-related validity of paper-and-pencil exams such as the National Teacher Examination (NTE) for teacher licensing and the lack of any apparent empirical linkages of the NTE to student outcomes of importance (e.g., scores on standardized achievement tests). Subsequently, they moved forward with the development of new, on-the-job teacher evaluation procedures using classroom-based observation instruments, most of which were touted as containing ways to measure teacher behaviors derived from findings in the process-product literature. The first state to move in this direction for the certification (licensing) of teachers was Georgia. A variety of other states, mostly in the Southeast, rapidly followed Georgia's lead. South Carolina, Florida, Mississippi, North Carolina, Tennessee, Virginia, Alabama, Arkansas, Texas, Kansas, Oklahoma, Connecticut, and many others are cases in point.

As these large-scale, statewide teacher evaluation efforts were implemented during the 1980s and on into the 1990s, a variety of issues pertaining to classroom-based assessments and evaluations of teachers and teaching began to emerge. Included were concerns for the following:

1. High- versus low-inference criteria and evaluative judgments
2. The length and number of observations needed to make reliable judgments
3. The role of the observer/evaluator in the classroom and the intrusiveness of observers
4. The number of teaching behaviors and scoring criteria necessary and sufficient to make valid judgments about competence
5. Who should evaluate
6. How many observations should be completed
7. The role of written lesson planning
8. The relationship between formative and summative elements of the evaluation process

9. Whether classroom visits should be announced, so that evaluations would reflect what teachers are capable of doing, or unannounced, so the observer could obtain an indication of typical practice ("can do" versus "is doing" arguments)
10. Criterion-referenced versus norm-referenced evaluation/decision-making models
11. Procedures for setting standards
12. How data should be aggregated to make summative evaluation decisions
13. The role of remediation and support for those failing to meet established standards
14. The effects of assessment demand characteristics on score inflation

Most of these issues have been discussed in detail in two comprehensive handbooks edited by Millman (1981) and Millman and Darling-Hammond (1990); space constraints do not allow their further elaboration here.

Classroom-based observation aimed at identifying characteristics of effective teachers had, by the middle 1980s, spilled over into large-scale, state-mandated programs to evaluate and license teachers based on what they actually do in classrooms, with primary attention given to those teaching behaviors identified through the process-product literature as reasonably related to student achievement and other student outcome variables (Brophy, 1986). In previous work, I have discussed the potential influence of these emerging large-scale teacher evaluation systems on increasing the professional legitimacy of principals as instructional leaders and supervisors, particularly from the perspective of principals' roles in assessing and assisting teachers to improve the quality of teaching and learning in classrooms (Ellett, 1987).

Of particular note and interest here, however, and given the arguments posed in this chapter, are the names of various large-scale teacher evaluation systems developed in many states as a result of the flurry of educational accountability activities in the 1980s. Metaphorically, these names speak to the primary evaluation concern—the teacher—largely through an evaluation focus on teacher behavior, with some variation in the role and degree of professional judgment (e.g., high- versus low-inference methodologies). For example, consider the following names of instruments developed in various states (some are still operational): the Teacher Performance Assessment Instruments (TPAI; Georgia), the Mississippi Teacher Assessment Instruments (MTAI; a version of the Georgia TPAI), the Assessments of Performance in Teaching (APT; South Carolina), the Florida (Teacher) Performance Measurement System (FPMS), the Virginia Teaching Practices Record (VTPR), the Tennessee Career Ladder Teacher Evaluation System (TCLTES), the Georgia Teacher Evaluation Process (GTEP), the Texas Teacher Appraisal System (TTAS), the Connecticut (Teacher) Competency Instrument (CCI), and, more recently (1991), the Louisiana Components of Effective Teaching (LCET). Clearly, and importantly, given the focus of the arguments in this chapter, the primary evaluation concern in each of these highly developed, classroom-based evaluation systems is the teacher and/or teaching—with an emphasis on the observation of teaching behaviors in the classroom.

It is also interesting to note that the major evaluation concern reflected in local school district evaluation systems typically mandated by local school board policy is also the teacher or teacher behaviors. The Teacher Assessment and Development System (TADS; Dade County, Florida) and the Teacher Assessment Process (TAP; Pasco County, Florida) are cases in point. Indeed, one would be hard-pressed to find an evaluation instrument or procedure used at the local district level that does not have the words *teacher evaluation* in the title. A national study of teacher evaluation practices in the 100 largest school districts completed in 1985 (Ellett & Garland, 1987) and a 10-year follow-up study recently completed (Loup, Garland, Ellett, & Rugutt, 1996) well document this fact. Furthering this point, past and present popular works in the field maintain titles and contents with a primary focus on the evaluation of the teacher (see, e.g., McLaughlin & Pfeifer, 1988; Medley et al., 1984; Millman, 1981; Millman & Darling-Hammond, 1990; Shinkfield & Stufflebeam, 1995).

What Is the Problem With Traditional Classroom-Based Assessment?

The central argument in this chapter is that there is a need to change both our current conceptions and our procedures for conducting

BEST COPY AVAILABLE

observation, assessment, and evaluation in classrooms. It seems apparent that the recent history of looking into classrooms, the emphasis given in classroom-based assessments to the findings of process-product and teacher effectiveness research, and the national and local district focus on the evaluation of teachers and/or teaching for educational accountability purposes have distorted the richness of classroom-based assessment and its potential to improve life in classrooms. These events have seemingly failed to recognize and reflect what I believe is most important in the implementation of classroom-based assessment—learning, for students, teachers, and principals as well. What is classroom-based assessment of teaching and learning and how does it differ from typical teacher evaluation methods and procedures?

How Can Classroom-Based Assessment Be More Productive?

Classroom-based assessment, as used here, refers to any data-gathering procedure or set of procedures used to collect information about actual teaching and learning processes in classrooms with three primary objectives in mind: (a) broadening our understanding of the complexities of teaching and learning, (b) enhancing the development of quality teaching and learning environments for both students and teachers, and (c) contributing to the knowledge base designed to link research, theory, and practice with the goal of enhancing learning for students, teachers, and perhaps administrators as well. Thus, the emphasis in developing and using this kind of assessment is consistent with models of formative rather than summative evaluation (Scriven, 1967). Assessment can be distinguished from evaluation by the fact that it is not used to make decisions about individuals (whether teachers or students) relative to established standards. From this perspective, assessment is an ongoing process and assessment data represent an ongoing set of "thermometers" that can be used to make inferences about the quality of teaching and learning and general life in classrooms. Evaluation, on the other hand, occurs when assessment data are compared with an established standard to make important summative decisions, as in annual teacher evaluations, certification and licensing of teachers, and so on.

A variety of assessment methods might be used to fulfill the three goals listed above. I provide a few examples later in this chapter. Whatever methods are used, however, they should be grounded in direct observation, personal experience, and perception of actual life in classrooms. Thus, student scores on standardized achievement tests and summative scores on teacher evaluation instruments (including portfolios) designed *only* for use in making employment, licensure, and certification decisions and the like are not examples of classroom-based assessment. The largest potential database for conducting such assessment consists of in situ observations accompanying existing policy-based teacher evaluation procedures. However, I take the position that, if they are to become true classroom-based assessments as defined here, these procedures must (a) be conceptually shifted to a learning-centered, rather than teacher- or teaching-centered, focus; (b) extend beyond the contributions of process-product research, the teacher effectiveness literature, and rigid perspectives grounded in philosophies of objectivism and behaviorism; and (c) be used in ways that broaden our understanding of the complexities of teaching and learning consistent with cognitive learning theories and more constructivist views of the nature of knowledge. Clearly, most existing teacher evaluation instruments and procedures are inadequate as classroom-based assessments, though many may have the potential to be used in this way.

Some Suggestions for Solutions

It seems to me that to fulfill the potential of classroom-based observation and assessment methodologies to improve life in classrooms, we need to do at least three things at both conceptual and operational levels. First and foremost, we need to understand that the most important focus in our work in classrooms should be the active process of student learning (not learning as inferred from standardized achievement test scores). A continued primary focus on observing, assessing, and evaluating teaching and/or the teacher and making inferences about students' learning from standardized test scores is not likely to result in major improvements in classrooms. The historical record seems quite clear on this point. The teacher evaluation focus of our past work in classrooms is also embedded in a large number of organizational and social context variables that suggest teacher evaluations over the years have been largely ritualistic and ineffective in improving classrooms (Darling-Hammond, Wise, & Pease, 1983; Scriven, 1981). The considerable difficulties in implementing and sustaining meaningful teacher evaluation efforts have also been recognized (McLaughlin, 1990). Thus, the yield of millions of classroom observations and standard evaluations of teaching and teachers to enhance teacher development and student learning has predictably been rather small.

We need to expand our observation and assessment focus in classrooms to teaching *and* learning, as well as to interactions between teacher and students and among students. In a recent document, I made a call for the development of a new generation of classroom-based assessments of teaching and learning (Ellett, 1990). Such assessments are grounded in a set of conceptual concerns and beliefs common to all classrooms (e.g., all students can learn, individual differences, student self-responsibility for learning, higher-order thinking skills, learning equity, etc.), and they require that the focus of our work in classrooms shift from assessing and evaluating sets of prescriptive teaching behaviors and the teacher to a primary concern for student learning. This approach recognizes that there is no "one best way" to teach, appreciates teachers' professional autonomy and choices, and is sensitive to variations in classroom life from one teacher to the next, from one day to the next, and to the wide range of social context variables that frame the quality of teaching, learning, and life in classrooms. Some (teachers particularly) have attacked this view, questioning the validity with which observers/assessors can make inferences about student learning. However, at the behavioral observation level there are many obvious indicators in classrooms from which inferences about student learning and the quantity and quality of cognitive engagement can be made (see Ellett, 1990, for some examples). Indeed, isn't it critical for teachers to actively observe, assess, and make inferences about their students' learning during the processes of teaching and learning as these are occurring? Good teachers do this as an ongoing process, and good observers/ assessors seemingly must do so as well.

Second, we need to attend to and appreciate emerging cognitive views of learning and the shifting curriculum emphasis on teaching students to think (Jones, Palinscar, Ogle, & Carr, 1987; Marzano et al., 1988). We also need to find ways through more comprehensive classroom-based assessments to integrate these ideas more fully with traditional behavioral observation and assessment in classrooms. The recent emphasis given to constructivist theories and methods of teaching and learning, particularly in math and science, should not go unrecognized either as we continue to observe and assess in classrooms (Taylor, Dawson, & Fraser, 1995; Tobin, 1993). There are a variety of perspectives on, and ontological and epistemological arguments about, the nature of constructivism and constructivist teaching and learning (Guba, 1990). These arguments aside, it seems clear that recording and assessing teacher and student behaviors, although useful, is not enough to enrich deeply our understanding of the complexities of teaching, learning, and classroom life. Probing our data further for the whats, hows, and whys of teaching and learning seems necessary. Philosophical arguments about whether we can ever fully understand another's interpretations of the world aside, finding ways to broaden and enhance our understanding of teachers' and students' perceptions, intentions, cognitions, and beliefs (meanings) can enrich our own understanding of classroom life and of teaching and learning processes.

Although the extant process-product research paradigm and research findings have undoubtedly made many contributions to our understanding of effective teaching, they do not have much to say about learning processes on the part of the student or the teacher. What seems needed in our future classroom-based assessments of teaching and learning is the extension of ideas about effective teaching to more important questions about teachers and students and the ways in which they learn, construct their understandings, and change. There are newer resources available to help us move in this direction, and some of these are described later in this chapter.

The third thing we need to do in developing a new generation of classroom-based assessments of teaching and learning is to attend to the value-addedness of using multiple methodologies in assessment. Direct, systematic observation probably remains the most widely used method of collecting information about classrooms. As detailed above, from the teacher evaluation perspective, few classroom-based measures (whether developed for use in state-mandated certification efforts or in local district evaluation programs) emphasize the collection of assessment data to make inferences about student learning as part of the assessment and/or evaluation process. The one exception I know of is the System for Teaching and Learning Assessment and Review (STAR; Ellett, Loup, & Chauvin, 1990). The brief discussion of the STAR that follows is included here to emphasize the need to refocus our assessment strategies as we work to improve classrooms, from the evaluation of teachers and teaching to student learning.

The System for Teaching and Learning Assessment and Review (STAR)

The STAR was originally developed from 1989 to 1991 in response to a state legislative mandate requiring the use of systematic classroom observation for the collection of data for the purpose of making teacher certification decisions. A more complete description of the genesis of the STAR as well as its conceptual foundation, structure, and associated assessment philosophy and procedures can be found in Ellett (1990). A summary of the conceptual and research base of the STAR as a comprehensive learning environment assessment can be found in Ellett et al. (1990). To reinforce points made in this chapter, the STAR is offered as an important and comprehensive assessment framework and process for looking into classrooms that requires the observer/assessor to judge the *effectiveness of teaching* by making inferences from observations of students' engagement in learning. Once systematic classroom observation data are collected, they are assessed (not evaluated) using a large number of assessment indicators that operationalize multiple Teaching and Learning Components (TLCs). A dichotomous assessment decision (acceptable or unacceptable) is made for each assessment indicator using the content/meaning of each indicator statement, classroom observation data collected/recorded, an accompanying annotation that thoroughly explains the meaning and provides examples of the assessment indicator, and a decision-making rule for the indicator. Thus, the STAR assessment process is deliberately designed to reflect and appreciate the importance of principals (and other assessors who make professional judgments about the quality of teaching *and* learning) giving strong consideration to the uniqueness of each classroom context in which teaching and learning occur.

The following list provides examples of some STAR assessment indicators; these indicators operationalize the Learning Equity TLC of the STAR:

1. Comments, questions, examples, demonstrations and/or other contributions are solicited from (or between) students throughout the lesson.
2. Students' contributions are considered and/or recognized.
3. Fairness and impartiality are evident in interactions with and among students.
4. Students are given reasons for actions, decisions, or directives as needed.
5. Functional elements of the learning environment are arranged to enhance the equity of learning activities.

The STAR TLCs and assessment indicators were developed to be consistent with findings from the process-product literature, but at the same time they were developed to extend this literature consistent with newer discussions about the importance of emphasizing teacher knowledge (Shulman, 1986) and expertise (Berliner, 1986) in teaching, learning, and the assessment process.

The assessment indicator represents an essential concept within the STAR structure and assessment process. An assessment indicator in the STAR is far more than a teaching behavior, and the STAR is not a teacher behavior checklist. Rather, it is a comprehensive assessment framework that is focused on student learning; it requires the assessor to make contextually sensitive, professional judgments about the quality of teaching *and* learning.

How is using a STAR assessment indicator different from evaluating a teacher using a more traditional teacher behavior checklist? In a teacher behavior checklist, an item might read, "The teacher's explanations of content and topics are clear." In the STAR, a similar assessment indicator might read, "Students clearly understand explanations of content and topics." As this example illustrates, the student-centered focus of indicator assessments in the STAR suggests at least the following possibilities, and there are obviously many more that reflect key differences between rating and evaluating teachers and teaching behaviors and making contextually based, professional assessments with indicators such as those reflected in the STAR:

1. What is of primary importance in classroom-based assessment is student learning and understanding, not what the teacher does or says.
2. Given the preceding point, the teacher does not have to say or do anything; students might arrive at clarity of understanding by other means (e.g., through cooperative learning in groups).
3. Clarity is defined in student learning terms, not by what the teacher says or does or by how clear the teacher's explanations are to the assessor.
4. Evidence for the clarity of students' understanding can be found in a variety of sources (e.g., student-to-student discussions, degree to which students can complete tasks without obvious errors due to misunderstanding and confusion, high quantity *and* quality of engagement, "aha" experiences and appearances, followed by successful task engagement).
5. The teacher behavior statement could be evaluated with no students in the class.

Thus, assessment indicators constituting the STAR are contextually bound, and a wide variety of observable conditions and classroom events must be considered when professional judgments are made about the acceptability or unacceptability of various indicators. No absolute evaluation standard to be used across teaching and learning contexts is available for making assessment decisions, nor is one implied by the assessment indicators. The assessment philosophy and process reflected in the STAR is nonprescriptive and allows for considerable teacher discretionary power about how best to teach. However, the various TLCs represent the kinds of general teaching and learning concerns in classrooms that have wide professional endorsement. Most would agree, for example, that managing time for learning is an important element of effective teaching and learning. However, the particular strategies that teachers and students might select and use to manage time, and the resultant observations of classroom events by an assessor, may be quite different in a large kindergarten classroom context as opposed to a small advanced physics classroom context in high school. The various STAR assessment indicators and assessment processes reflect concern for the differing contexts from one classroom to the next and from day to day in a single classroom. This approach might be contrasted with the great majority of teacher evaluation instruments and procedures, which require evaluators to make performance ratings of particular valued teacher characteristics or teaching behaviors.

Though designed to refocus and extend classroom-based observation and assessment systems to encompass student learning, the STAR is not sufficient to do the kinds of in-depth, classroom-based assessments proposed here, which are needed to make stronger inferences about what teachers and students know and understand, how they derive and construct meaning from their classroom experiences, what they believe about themselves as learners (including teachers as learners), how they learn, what their preferences and bases of motivation for learning are, and so on. At the same time, we cannot measure how teaching and learning proceed through human attestations alone. Learning occurs within both internal (personal, constructed) and external (interactive) environments. What kinds of assessment procedures might be used to enhance systematic classroom observation procedures such as the STAR?

A Call for Multiple Methodologies

There are a variety of methods and resources available that can be integrated with direct observation systems to extend our understanding of teaching and learning in classrooms. Some of these procedures are more formal than others, but all have great possibilities when integrated in various combinations with systematic classroom observation methods. For example, the traditional clinical supervision model of preobservation conference, observation, postobservation conference still has possibilities to broaden our understanding of teaching and learning and possibilities for the professional development of teachers and assessors as well. In my past assessment work with teachers, and when I teach a graduate course in the supervision of instruction, I have used an expanded supervision model that includes the following assessment elements:

1. Development by the teacher of a comprehensive unit plan for a 5- to 7-day unit of teaching and learning for a class of the teacher's own choosing. (I always want to see teachers at their best, and I believe in giving them choice in assessments.)
2. Review of the plan and associated teaching and learning activities and materials, followed by a semistructured interview/conference designed to enable the evaluator to do the following:
   (a) to make inferences about what the teacher knows and understands about individual differences among students, the form and structure of content knowledge, the nature of how students learn, the rationale (in terms of student learning) for planning various teaching and learning activities, informal assessments (of student learning) planned, the use of aids and materials in a manner that enhances student learning, home learning assignments planned, and so on; and
   (b) to establish assessment context factors by means of a thorough description of the nature of learners in the class to be observed (e.g., their motivation levels, typical interests, family backgrounds, special needs)
3. A preobservation discussion with the teacher immediately preceding the observation of teaching and learning to frame the context of the observation further and to attend to any special considerations that need to be made (e.g., children who are not feeling well, unavailable resources, classroom environment conditions, or anything else the teacher believes is helpful in developing my understanding of the teaching and learning context)
4. Direct, systematic observation of teaching and learning, with a strong focus on students, for a minimum of a whole class period for any one observation, seated where I can easily see students and teaching and learning activities (not at the back of the classroom); taking necessary and sufficient notes through a balance of note taking and observing—*not total lesson scripting*—regularly scanning the entire class at fixed intervals to build a class index of *both* the quantity and quality of student engagement (task and cognitive engagement) in learning

5. Postobservation discussion with the teacher immediately after the observation to clarify (for me) any unusual teaching and learning events or activities observed
6. A personal period of quiet reflection and consideration of the observation data recorded in my notes, adding fine-brush details from immediate memories (usually 20-30 minutes)
7. A rough synthesis and recording of teaching and learning highlights from the classroom observation
8. A considerable length of time spent with the STAR using the various assessment indicators as a guide for making professional judgments about the quality of teaching and learning observed (usually 30-45 minutes)
9. Preparation for a postobservation conference with the teacher—developing questions of interest, developing hypotheses to be tested further, and so on
10. A postobservation conference with the teacher to share observations (from this assessment or from the many other *movies in the teacher's head* derived from a variety of other lessons/teaching and learning experiences that only the teacher has), test hypotheses, collect further information, develop a better understanding of how teaching and learning are perceived and constructed (by the teacher, the students, and me), share ideas for enhancing the quality of teaching and learning, plan future assessment activities, and so on

Over the years, I believe this expanded clinical supervision model and assessment procedure has worked well for me. However, there are a variety of other possibilities for adapting/adopting this model and for broadening classroom-based assessments of teaching and learning. Other suggested methods include, but certainly are not limited to, the following:

1. The use of highly developed student and teacher perception measures of characteristics of the classroom or school learning environment (both class and personal, constructivist-based learning forms are now available)
2. Random selection of students from classes observed to collect interview data to be integrated with teacher interview and direct observation data

3. Development of teacher and student portfolios over time, combined with periodic interviews
4. Use of student and teacher *self-reflection learning diaries*
5. Measures of student and teacher learning efficacy and beliefs about learning
6. Joint interviews with students and the teacher to share and interpret classroom observations of teaching and learning
7. Videotape analysis by and/or with the teacher and/or students
8. *Critical events in learning discussions* among members of the class

Each of these methods and a variety of others can be used to enhance and enrich our understanding of teaching and learning, and larger life in classrooms. Given the complexities of teaching and learning, the great variety of individual differences among teachers and learners, and the current value being placed on cognitive learning theories and constructivist views of knowledge, none of these methods seems sufficient when used alone, but all are rich possibilities. The recent work of current classroom researchers in math and science education, such as Ken Tobin at Florida State University, Campbell McRobbie at Queensland University of Technology in Australia, Michael Roth of the University of Victoria, British Columbia, Barry Fraser and Peter Taylor at Curtin University in Australia, provides examples of how the varieties of classroom-based assessment listed above can extend beyond contributions of the process-product literature and move us toward greater understanding of teaching and learning *processes* that better incorporate cognitive and constructivist views of teaching and learning.

Synthesis and Projection

In this chapter I have argued that classroom-based assessments of teaching and learning can be enriched by the incorporation of assessment philosophies and multiple methodologies that extend beyond those derived from earlier process-product literature and teacher effectiveness research. At issue with the vast majority of direct classroom observation systems emanating from past educational accountability and summative, personnel evaluation concerns, has been a predominant focus on the teacher and teaching, rather than on student learning. I have made three suggestions here for changing our conceptual and procedural focus as we move forward to develop better classroom-based assessments of teaching and learning: (a) Focus on student learning, (b) attend to and accommodate newer theories of cognitive learning and constructivist views of knowledge, and (c) integrate multiple methodologies in our assessment strategies.

Classroom-based assessments remain at the core of inferences we make as professionals about the nature and quality of teaching and learning—whether made by teachers, researchers, school principals, or others. If we believe that teaching and learning are to be understood as active behavioral and cognitive processes, with their meanings individually constructed by teachers and students, our future assessments of teaching and learning must include more than standardized teacher evaluation instruments and standardized student achievement tests. Over the years, I suspect we really haven't learned much about teaching and learning *processes* from the millions of scores generated by these kinds of measures. In addition, our future research must appreciate and build upon, but considerably extend, the process-product and teacher effectiveness literatures. We are participating in an exciting time in the development of newer forms of classroom-based assessments, and it appears that future research, theory development, and practice based on these assessments can contribute much to our understanding of the complexities of teaching and learning and life in classrooms.

Consistent with the views of the National Board of Professional Teaching Standards, I also believe that the quality and breadth of our classroom-based assessments of teaching and learning can do much to enhance the professionalization of teaching (and administration) —not from the narrow social/political point of view reflecting standardized teacher evaluation practices and increased educational accountability, but from the rich opportunities that such assessments provide to enhance the inquiry skills of teachers, students, administrators, and others, and to construct new insights and understanding about teaching and learning. Using newer forms and combinations of assessments in creative ways should become a primary goal of principals, teachers, and others as we attempt to enhance the quality of teaching and learning in our schools in the 21st century. This chapter has offered some reflections on the role of comprehensive, classroom-based assessments of teaching and learning with that goal in mind.

References

Beecher, D. E. (1949). *The evaluation of teaching: Backgrounds and concepts*. New York: Syracuse University Press.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*(7), 5-13.

Brophy, J. (1986, October). Teacher influences on student achievement. *American Psychologist, 41*(10), pp. 1069-1077.

Brophy, J. (1988). Research on teacher effects: Uses and abuses. *Elementary School Journal, 89*, 3-22.

Charters, W. W., & Waples, D. (1929). *The Commonwealth Teacher Training Study*. Chicago: University of Chicago Press.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53*, 285-328.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Ellett, C. D. (1987). Emerging teacher performance assessment practices: Implications for the instructional supervision role of school principals. In W. Greenfield (Ed.), *Instructional leadership: Concepts and controversies* (pp. 302-327). Boston: Allyn & Bacon.

Ellett, C. D. (1990). *A new generation of classroom-based assessments of teaching and learning: Concepts, issues and controversies from pilots of the Louisiana System for Teaching and learning Assessment and Review* (Project document, Louisiana Statewide Teacher Evaluation and Teaching Internship Projects). Baton Rouge: Louisiana State University, College of Education.

Ellett, C. D., & Garland, J. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to "state-of-the-art" systems? *Journal of Personnel Evaluation in Education, 1*, 69-92.

Ellett, C. D., Loup, K. S., & Chauvin, S. W. (1990). *The System for Teaching and learning Assessment and Review (STAR)* (Assessment manual, Louisiana Statewide Teacher Evaluation and Teaching Internship Projects). Baton Rouge: Louisiana State University, College of Education.

Flanders, N. A. (1970). *Analyzing teaching behavior*. Reading, MA: Addison-Wesley.

Gage, N. (1972). Can science contribute to the art of teaching? In *Teacher effectiveness and teacher education: The search for a scientific basis* (pp. 27-39). Palo Alto, CA: Pacific.

Gage, N., & Needels, M. (1989). Process-product research on teaching: A review of criticisms. *Elementary School Journal, 89*, 253-300.

Good, T. L., & Mulryan, C. (1990). Teacher ratings: A call for teacher control and self-evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 191-215). Newbury Park, CA: Sage.

Guba, E. G. (Ed.). (1990). *The paradigm dialog*. Newbury Park, CA: Sage.

Jones, B. F., Palinscar, A. S., Ogle, D. S., & Carr, E. G. (Eds.). (1987). *Strategic teaching and learning: Cognitive instruction in the content areas*. Alexandria, VA: Association for Supervision and Curriculum Development.

Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education, 10*(3), 203-226.

Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

McLaughlin, M. W. (1990). Embracing contraries: Implementing and sustaining teacher evaluation. In J. Millman (Ed.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 403-415). Newbury Park, CA: Sage.

McLaughlin, M. W., & Pfeifer, R. S. (1988). *Teacher evaluation: Improvement, accountability, and effective learning*. New York: Teachers College Press.

Medley, D. M. (1977). *Teacher competence and teacher effectiveness: A review of process-product research*. Washington, DC: American Association of Colleges of Teacher Education.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.

Medley, D. M., & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In N. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Millman, J. (Ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.

Millman, J., & Darling-Hammond, L. (Eds.). (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Sage.

Nutt, H. (1928). *Current problems in the supervision of instruction*. New York: Johnson.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership, 45*(8), 74-85.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Skokie, IL: Rand McNally.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 244-271). Beverly Hills, CA: Sage.

Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Hingham, MA: Kluwer Academic.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Simon, A., & Boyer, E. C. (1967). *Mirrors for behavior: An anthology of classroom observation instruments* (6 vols.). Philadelphia: Research for Better Schools.

Stallings, J. (1977). *Learning to look: A handbook for classroom observation and teaching models*. Belmont, CA: Wadsworth.

Taylor, F. (1947). *Scientific management*. New York: Harper.

Taylor, P., Dawson, V., & Fraser, B. (1995). *Classroom learning environments under transformation: A constructivist perspective*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Tobin, K. (Ed.) (1993). *The practice of constructivism in science education*. Washington, DC: American Association for the Advancement of Science.

6

Client Surveys in Teacher Evaluation

JAMES H. STRONGE
LAURA P. OSTRANDER

Teachers play an essential role in the success of schools and schooling. Research supports the premise that teachers are among the most powerful determinants of student learning (see, for example, Goodlad, 1984; Northwest Regional Educational Laboratory, 1990). Moreover, budget allocations for teacher salaries, the size of the teaching force, and dependence on teachers as the primary workforce of schools point to teachers as central figures in school success. For this reason alone, the evaluation of teaching deserves high priority (Harris, 1986).

Teacher evaluation is a critical factor in any effort to validate teaching and learning and the success of schools. Although teacher evaluation has gained attention in recent years, it remains a neglected area of educational research. Evaluation practices have been labeled

---

AUTHORS' NOTE: We would like to express our appreciation to the Lenawee (Michigan) Intermediate School District, the Williamsburg-James City County (Virginia) Public Schools, and the Virginia Beach City (Virginia) Public Schools for permission to reproduce their client survey forms in this chapter.

seriously deficient (Haefele, 1992), chaotic (Medley, Coker, & Soar, 1984), and a disgrace (Scriven, 1981). Regarding the state of the art of teacher evaluation, Frase and Streshly (1994) summarize the opinions of several writers when they state, "Research and learned opinion strongly support the contention that teacher evaluation has been of little value" (p. 48).

In an era when accountability is the mantra, educational policy makers, educational leaders, legislators, and the American public in general view improving teacher assessment as an important step toward ensuring educational quality. Consistent with the contributions of other chapter authors in this text, we provide here a rationale for seeking new approaches to teacher evaluation and explore some potential benefits of collecting data from parents, students, and peers for the assessment of teacher proficiency and for improving classroom instruction and other areas of teacher performance. Specifically, we explore the following four questions:

1. Why should client data be used in teacher evaluation?
2. What are the benefits of including parents in the teacher evaluation process?
3. Can students provide reliable and useful measures of teaching effectiveness?
4. How can feedback from peers be used most effectively in teacher evaluation?

Why Should Client Data Be Used in Teacher Evaluation?

The teacher evaluation methods used in the majority of school districts today are based on a model that requires administrators to diagnose weaknesses and subsequently to prescribe solutions. In many American school districts, the primary teacher assessment tool is classroom observation, which may or may not be announced in advanced (Educational Research Service, 1988). Although this single-source method of assessing teacher performance has serious shortcomings (McGreal, 1994; Medley et al., 1984; Scriven, 1981), it is firmly entrenched in American schools.

Medley et al. (1984) relate the condition of teacher evaluation to continued reliance on a practice filled with opportunities for human error: "The whole art of teacher evaluation up to the present consists of obtaining someone's subjective judgment of how good a teacher is, a judgment based on the assumption that the judge knows what good teaching is and can recognize it when he sees it" (p. 4). Furthermore, there is considerable disagreement as to what constitutes the best practices with regard to the complex act of teaching (Scriven, 1981; Tetenbaum & Mulkeen, 1988).

This subjective-judgmental teacher evaluation model, which makes *observation* virtually synonymous with *evaluation*, has changed little in the past century (Stronge & Helm, 1991). Prevailing practices do not reflect research in effective teaching, nor do they support ongoing educational reform efforts or restructured environments (McGreal, 1994; Peterson, 1995). The most serious indictment against business as usual in assessing the job performance of classroom teachers is that the process has done little to improve instruction (Frase & Streshly, 1994).

Classroom Observation Does Not Equal Good Evaluation

Classroom observations are conducted on the premise that seeing a teacher in action is the best way to gather data for judging that teacher's effectiveness. Classroom visits, although typically narrow in scope, are important in that they allow evaluators to assess classroom climate, observe teacher-student interactions, observe instruction directly, and see classroom functioning as no other strategy can (Evertson & Holley, 1981). Direct classroom observation can be a useful way to collect information on teacher performance; as a stand-alone data collection process, however, it has major limitations. If the purpose of a teacher evaluation system is to examine a comprehensive picture of performance in order to guide professional growth, then classroom observation furnishes only one piece of the information collected.

Scriven (1981) provides an explanation for the low reliability of ratings made during classroom visits. First, the number and length of observations are almost always inadequate for making generalizations. Second, evaluators focus attention on their own personal interests; thus, what they notice reflects their personal viewpoints. Third, poor recording systems force observers to rely on recollections that are influenced by preexisting conceptions. Fourth, any personal relationships or alliances between evaluators and their subject present confounding factors. Finally, the visit itself alters the behaviors of teacher and students, narrowing the chances of the evaluator's seeing a representative sample of teaching—a factor that pre-announced visits only exacerbate.

Additional concerns regarding classroom observation include the following (Stronge, Helm, & Tucker, 1995):

- The artificial nature of scheduled observations
- The limited focus of any given observation
- The infrequency of observations (e.g., two visits every 2 years for classroom teachers)
- The fact that only a portion of the full repertoire of teacher duties and responsibilities can be observed in any one observation (e.g., selected teacher responsibilities may not be performed during the classroom visit)
- The fact that other teacher responsibilities occur entirely outside of the classroom

Given such concerns, the fallacy of making classroom observation the sole source of data to be used in assessing teachers is clear.

Although classroom observation can be a meaningful and vital aspect of a comprehensive teacher evaluation system, it has major drawbacks as a single-source methodology. Teachers often are rated in terms of their perceived performance rather than their actual performance, and observers' perceptions can be affected by a variety of factors unrelated to the quality of instruction (Frase & Streshly, 1994; Wood, 1992). There is, in fact, little agreement among evaluators on the assessment of classroom performance. Like beauty, instructional quality seems to be in the eye of the beholder. All too frequently, whim and caprice color judgments (Haefele, 1993), and, although unintentional, prejudice or bias can be factors as well (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). Also, it is difficult to separate a teacher's value to the organization from responsibilities directly linked to classroom effectiveness (Scriven, 1981).

Despite continued reliance on observation data from principals, many teachers do not perceive principals to be competent evaluators of classroom performance, and there is considerable evidence to indicate that principals' ratings tend to be inflated (Frase & Streshly, 1994). Peterson and Kauchak (1982) blame the failure to develop more effective evaluation procedures on the state of the art, the need to satisfy diverse audiences who compete for information, and the fact that other data sources have been neglected.

Frase and Streshly (1994) have reported on a study involving six school districts in which paid auditors reviewed tenured teachers' evaluation data and made classroom observations. The auditors described many incidents of poor instructional practices (e.g., exclusive use of seatwork, extensive copying of exercises from chalkboards or overheads, students left unattended by their teachers) in classrooms where instructors received high ratings. They concluded that the principals' ratings were grossly inflated and offered little helpful information for teacher improvement.

Root and Overly (1990) have reached similar conclusions. They describe teacher evaluation as "a bureaucratic requirement that is conducted perfunctorily and does little to improve teacher performance" (p. 35). All too frequently, teachers do not receive constructive information from administrators, and they view the process as useless. Frase and Streshly (1994) describe the principal's performance in teacher evaluation as an "abrogation of duty [that is] malpractice in its worst form" (p. 55).

Bridges (1986) speculates that one explanation for the high ratings principals commonly give teachers is that they are unable to devote enough time to do the job adequately and thus are reluctant to be critical of teachers' performance. Many are defensive about the whole issue of performance assessment. Furthermore, some principals may want to avoid conflict, may fear repercussions, and may wish to avoid the high cost of litigation if they are challenged in court.

Even more serious criticisms of heavy or sole reliance on one-shot observations for teacher evaluation are the contentions that teachers seldom use evaluation data from principals to make meaningful changes in classroom practice (Frase & Streshly, 1994) and that current evaluation procedures protect less competent staff members from dismissal (Haefele, 1993). Likewise, there is evidence that principals seldom utilize evaluation data for personnel matters (e.g., hiring, firing, promoting) (McGreal, 1994).

There are two commonly stated arguments against continued reliance on formal observation as the sole assessment of teaching quality: First, the practice places the teacher's fate in the hands of a single judge (Harris, 1986; Peterson, 1995); and second, it is a model that no longer reflects the way American schools operate (Ellett & Garland, 1987; McGreal, 1994). Recent reform movements that foster stakeholder empowerment, collaboration, collegiality, and increased teacher accountability require better ways to measure teacher effectiveness (Dixon, 1994).

Collaboration in Teacher Evaluation

One reason for the uncertainty surrounding current evaluation methods involves the relative isolation in which teachers work. Bruce Joyce (cited in Costa, 1993) has described teaching as one of the most private acts in which humans engage, and Peterson (1995) cites isolation as a major sociological constraint on evaluation practices. Fullan (1994) has called for a move from isolation to collaboration for implementing the successful change strategies that current reform movements demand. He maintains that if organizations do not collaborate externally as well as internally, change does not occur.

One contemporary school reform movement that has significant implications for collaboration in teacher evaluation is total quality management (TQM) (Deming, 1986). Adopted by many school districts as a framework for educational reform, TQM offers techniques for assessing and improving all components of school operations (Bayless et al., 1992; Bonstingl, 1992). The application of TQM principles to education parallels a growing interest in applying successful strategies from the business world to the management of schools. Two elements drawn from an emphasis on organizational quality significant to the evaluation of teaching are pride of workmanship and customer service.

Pride of workmanship places the responsibility for service improvement on the supplier and requires constant monitoring through feedback from customers. Using this model, teachers are the service providers (or suppliers) and the customers (or clients) include students, parents, other teachers, other institutions, and the community. Client surveys are the primary source of customer feedback.

As more school districts have embraced reform efforts that focus on quality improvement, a concern for *customer service* and client

TABLE 6.1 Clients as Data Sources for Teacher Evaluation, by School Level (in percentages)

Data Source	*Elementary School*	*Middle School/ Junior High*	*High School*
Principals	95.3	93.4	90.8
Assistant principals	43.8	67.3	71.3
Department chairs	5.7	14.1	23.5
Supervisors	30.4	29.9	29.9
Peers	6.4	6.2	6.2
Students	2.8	3.0	3.4
Parents	1.2	1.1	1.1
Self-assessment	23.2	22.6	21.7

SOURCE: Adapted from Educational Research Service (1988).

satisfaction has emerged (Bonstingl, 1992). Teachers have identified both internal clients (e.g., principals, students, other teachers, department chairpersons) and external clients (e.g., parents, institutions of higher education, employers). Feedback information from one or more of teachers' clients can help fill a serious gap in the performance review arena. A summary of the use of client feedback in teacher evaluation is provided in Table 6.1.

The following sections of the chapter explore the issues raised by the incorporation of data from three key client sources in a comprehensive teacher evaluation system: parents, students, and peers.

What Are the Benefits of Including Parents in the Teacher Evaluation Process?

Numerous educational reform efforts of the past two decades have called for increased parental involvement in education and increased empowerment of parents for decision making. There is substantial evidence that good parent-teacher relationships and effective home-school communications have positive effects on school success (Becher, 1984). Traditionally, however, most teacher-initiated interactions with parents result from problems that occur at school (Mager, 1980). Some responsibility for this state of affairs can be attributed to educators who have become gun-shy because of experience with groups who oppose specific causes (e.g., sex education or outcome-based education), but the failure of parents to demand school accountability is undoubtedly a factor as well.

One vital way to enhance parental involvement in education is to invite parents into the educational enterprise as partners in decision making. In a discussion of the value of involving stakeholders (e.g., parents) in evaluation, Mark and Shotland (1985) note that the benefits include empowerment. Once invited to participate in the teacher assessment process, parents can offer perspectives and insights that are otherwise not available.

Epstein (1985) undertook one of the more comprehensive efforts to date to investigate the potential contributions of parents to teacher evaluation. In a study involving 77 elementary school teachers, Epstein compared and analyzed ratings from parents and principals to isolate variables that influenced judgments about teacher merit and to assess the efficacy of parent involvement in the evaluation process. The study revealed that the principals and parents emphasized different components of teaching. For example, principals were more aware of how well teachers performed extra duties than of changes in classroom practice; thus, doing a good job on noninstructional tasks may have resulted in higher marks from principals than exemplary teaching. Parents, however, were more knowledgeable about special efforts teachers made to help their children, and this insight had an impact on parents' ratings of teacher performance. Epstein maintains that parents may, in fact, be more cognizant than principals of some aspects of teacher quality:

> [The parents] may be knowledgeable about how the teacher interacts with the child and family, responds to the student's needs and skills, assigns appropriate challenges in books and in homework, and inspires the student to continue commitment to school work at home—all indicators of effective teaching. With unique interests and investments in both teacher effectiveness and school organization, parents may be legitimate and important contributors among multiple judges in the evaluation of teachers and school programs. (p. 4)

One of the most important variables Epstein identifies involves the teacher's interpersonal skills. Increased parental involvement, more teacher-initiated communication with parents, and good classroom discipline earned higher ratings for the teachers in this study. Epstein suggests that inclusion of parents in the evaluation process would result in "more teachers [who are] recognized as good teachers for different skills and teaching abilities than if only principals rated the teachers" (p. 9).

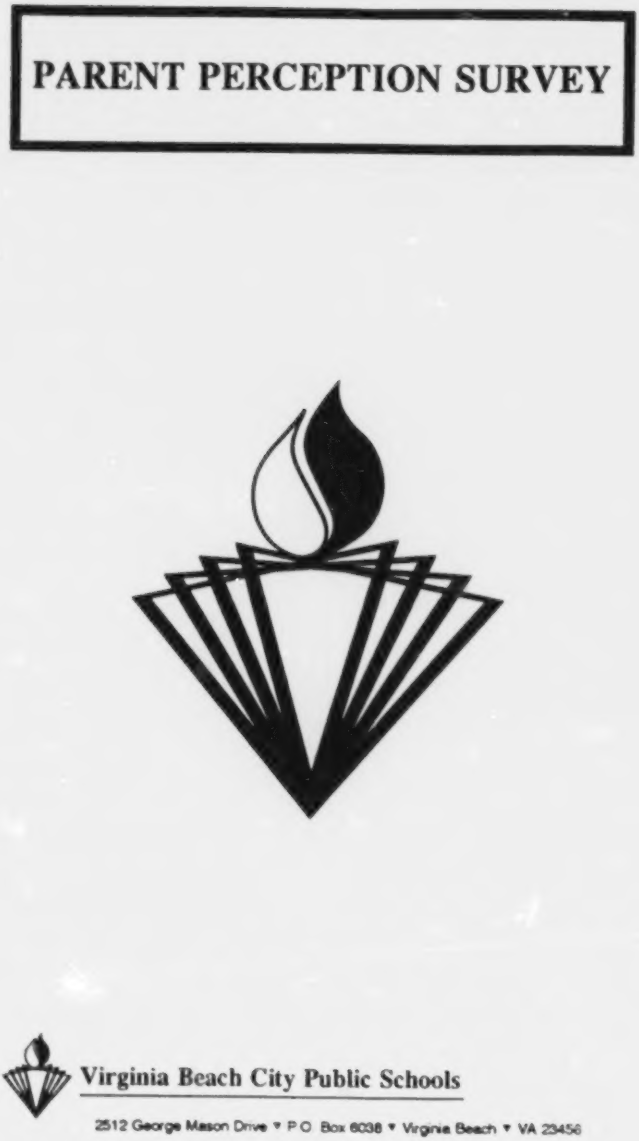
Peterson (1988) examined the use of parent surveys as an optional data source and found that 48% of the eligible teachers in two school districts (*n* = 701) chose parent surveys as a source of data for merit pay consideration. The survey they used was research based, but "addressed only information for which the parents could be expected to have first-hand information" (p. 243). The unit of analysis for this study was the class mean on one global item assessing parents' overall satisfaction with their children's classroom experience. Teachers received a mean rating of 4.49 on a 5-point scale. Both the teachers' mean rating and the return rate declined at higher grade levels. In follow-up interviews, 76.6% of the participating teachers described the parent surveys as providing "valuable feedback" that was "pertinent, specific, and ultimately credible" (p. 247). They suggested that the parents' open-ended comments were especially helpful.

Peterson's (1988) findings corroborate Epstein's work with regard to ratings by parents of elementary students, but they leave questions about the use of parent ratings for secondary teachers. He suggests that "less contact and communication results in more global or halo ratings for the teachers of older students" (p. 247).

Faucette, Ball, and Ostrander (1995) have reported on one of the most extensive uses of client surveys in teacher assessment in a large suburban school district (*n* = 75,000) in which teachers were encouraged to participate voluntarily in a customer service project. For 3 years, an instrument called the Parent Perception Survey, which was developed by teachers, was mailed directly to the parents of students of the participants and returned to an external agency by mail (see Figure 6.1). Confidential data summaries were provided for the teachers that were intended for personal and professional improvement.

The results were consistently positive for individual participants and for the school district as a whole. More than 70% of all teachers

(text continues on page 142)



PARENT PERCEPTION SURVEY

Virginia Beach City Public Schools

2512 George Mason Drive ▾ P.O. Box 6038 ▾ Virginia Beach ▾ VA 23456

Figure 6.1. The Parent Perception Survey
SOURCE: Virginia Beach City Public Schools, copyright March 1995. Reprinted by permission.

We welcome your comments here.
These comments will be given to your child's teacher.

Thank you for your cooperation and your assistance. You have made a valuable contribution to the education of the children of Virginia Beach. If you would like to discuss this further, please call (804) 427-4381.

Alternative formats of this publication which may include taped, braille, or large print materials are available upon request for individuals with disabilities. Call or write Jody Blaufas, Office of Public Information, Virginia Beach City Public Schools, 2512 George Mason Drive, P.O. Box 6038, Virginia Beach, Virginia 23456-0038. Telephone: (804) 427-4320 (Voice); 427-8201 (TDD). March 1995.

Teacher:

Grade or Subject:

School:

The teacher named above has volunteered to take part in the Parent Perception Survey. This teacher values your opinion. Please be sure your responses apply only to this teacher.

Start Here:

Since the **beginning** of the school year, have you...

- Attended a general school meeting, for example, back to school night or a meeting of a parent-teacher organization? Yes ☐ No ☐
- Attended a school or class event such as a play, sports event, or science fair? Yes ☐ No ☐
- Acted as a volunteer at the school or served on a school committee? Yes ☐ No ☐

INSTRUCTIONS

Please answer the three questions above. Then, carefully consider each of the following statements as it relates to your impressions about your child's teacher. Decide if you STRONGLY AGREE, AGREE, DISAGREE, or STRONGLY DISAGREE with the statement and fill in the appropriate block. If you are unable to make a judgement for any reason, mark the NO ANSWER BLOCK. Please do not leave any items blank.

If you have additional comments, please use the space on the back of this survey. To ensure the confidentiality of your response, tear off the section of the survey with your name and address. Then, please fold and return the completed survey to Continental Research by using the enclosed envelope. Please respond within one week.

Thank you very much for your time.

← TEAR OFF BEFORE MAILING

Figure 6.1. Continued

My child's teacher:	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
1) Communicates with me	☐	☐	☐	☐	☐
2) Makes me feel comfortable contacting him/her	☐	☐	☐	☐	☐
3) Provides helpful information during conferences	☐	☐	☐	☐	☐
4) Informs me about my child's progress	☐	☐	☐	☐	☐
5) Is courteous	☐	☐	☐	☐	☐
6) Listens to what I have to say	☐	☐	☐	☐	☐
7) Has my respect	☐	☐	☐	☐	☐
8) Informs me about classroom expectations	☐	☐	☐	☐	☐
9) Informs me of classroom grading procedures	☐	☐	☐	☐	☐
10) Makes classroom rules known to students	☐	☐	☐	☐	☐
11) Uses fair disciplinary procedures	☐	☐	☐	☐	☐
12) Challenges my child	☐	☐	☐	☐	☐
13) Builds on my child's strengths	☐	☐	☐	☐	☐
14) Recognizes improved achievement	☐	☐	☐	☐	☐
15) Recognizes positive behaviors	☐	☐	☐	☐	☐
16) Is respected by my child	☐	☐	☐	☐	☐
17) Fosters growth in areas needing improvement	☐	☐	☐	☐	☐
18) Encourages classroom participation	☐	☐	☐	☐	☐
19) Uses a variety of instructional methods	☐	☐	☐	☐	☐
20) Grades my child's overall classroom performance	☐	☐	☐	☐	☐
21) Relates instruction to everyday life	☐	☐	☐	☐	☐
22) Assigns homework that supports classroom learning	☐	☐	☐	☐	☐
23) Explains the homework assignment	☐	☐	☐	☐	☐
24) Offers additional help when necessary	☐	☐	☐	☐	☐
25) Provides an inviting environment	☐	☐	☐	☐	☐
26) Creates a safe learning environment	☐	☐	☐	☐	☐
27) Provides an environment that promotes learning	☐	☐	☐	☐	☐

TABLE 6.2 1993-1995 Parent Perception Survey Data for All Participating Teachers

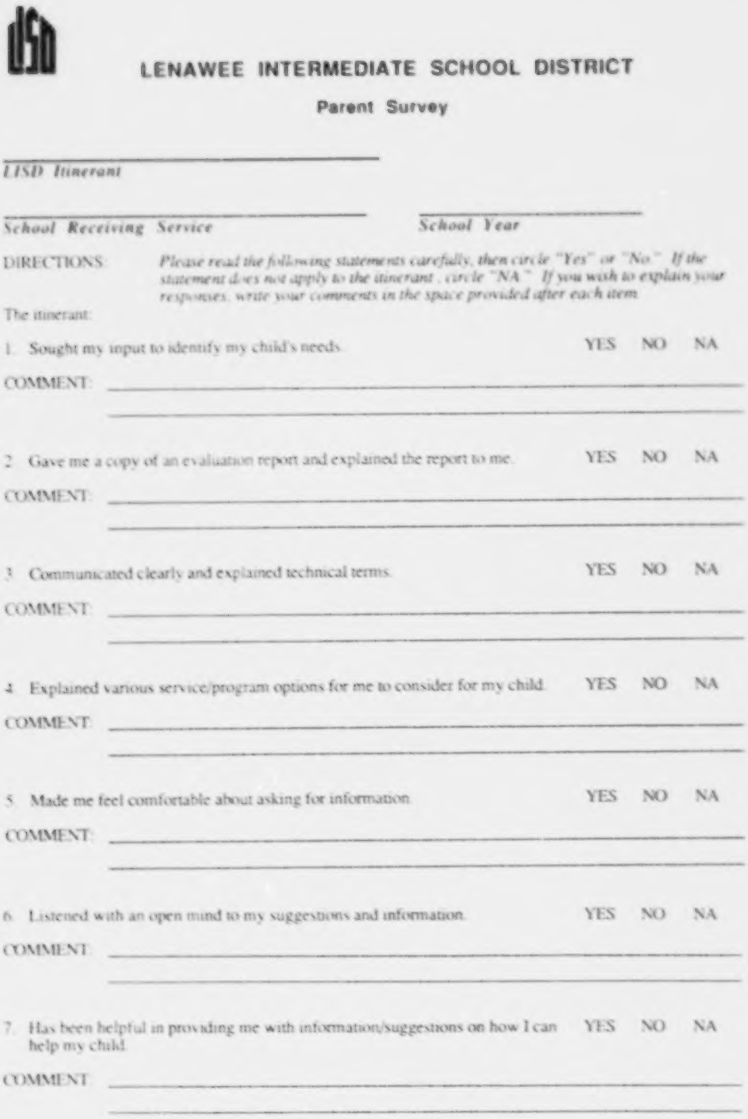
Category	*Teacher Participants* *No.*	*%*	*Range of Item Means*	*Number of Surveys Mailed*	*Response Rate (%)*	*Grand Mean*
1993	1,610	37	1.8-3.6	40,128	NA	3.40
1994	2,380	52	3.3-3.6	69,102	39.3	3.46
1995	1,740	38	3.4-3.6	51,972	31.4	3.53

SOURCE: Ostrander (1995).

in the district utilized the survey at least once, and many volunteered all 3 years. Each year the grand mean increased slightly (see Table 6.2), suggesting that teachers used the information they received from parents to improve the services they provided.[1] The highest impact appeared to be in the area of communication. Teachers reported that the primary modifications they made in their teaching practices (in response to the parent feedback) were increases in both the quantity and the quality of their communication with the parents of their students.

In a study involving a sample (n = 93) of the parent survey participants, Ostrander (1995) compared the ratings of the parents with teacher self-assessments and with students' and principals' ratings of the teachers on the items in the Parent Perception Survey. Data were disaggregated for six subcategories: classroom environment, grading, homework, communication, instruction, and interpersonal relationships. All ratings were generally quite high, but there were some subtle differences. Principals consistently rated the teachers higher than any of the other respondents on all subcategories except homework, where the teacher self-assessments were higher. The students gave the teachers the lowest ratings in all areas; the parents were second lowest. Additionally, parent ratings showed a higher correlation with student ratings than with teacher or principal ratings.

The notion that parents favor being involved in the evaluation process and that they appreciate being asked to provide feedback concerning teacher effectiveness has been validated in several projects (e.g., Faucette, Ball, & Ostrander, 1994, 1995; Peterson & Mitchell,



LENAWEE INTERMEDIATE SCHOOL DISTRICT

Parent Survey

LISD Itinerant

School Receiving Service *School Year*

DIRECTIONS: *Please read the following statements carefully, then circle "Yes" or "No." If the statement does not apply to the itinerant, circle "NA." If you wish to explain your responses, write your comments in the space provided after each item.*

The itinerant:

1. Sought my input to identify my child's needs. YES NO NA

COMMENT:

2. Gave me a copy of an evaluation report and explained the report to me. YES NO NA

COMMENT:

3. Communicated clearly and explained technical terms. YES NO NA

COMMENT:

4. Explained various service/program options for me to consider for my child. YES NO NA

COMMENT:

5. Made me feel comfortable about asking for information. YES NO NA

COMMENT:

6. Listened with an open mind to my suggestions and information. YES NO NA

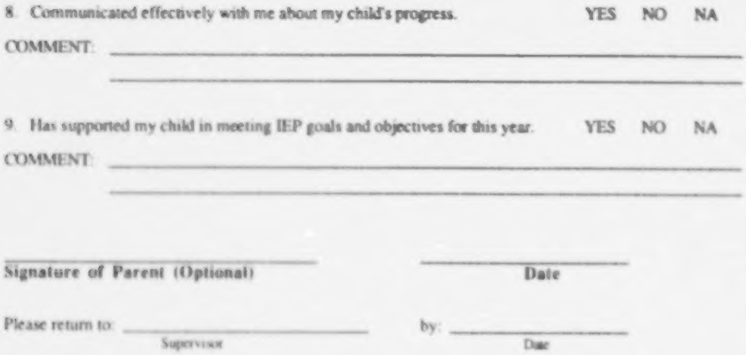
COMMENT:

7. Has been helpful in providing me with information/suggestions on how I can help my child. YES NO NA

COMMENT:

Figure 6.2. Sample of a Parent Survey
SOURCE: Lenawee Intermediate School District. Reprinted by permission.

BEST COPY AVAILABLE



8. Communicated effectively with me about my child's progress. YES NO NA

COMMENT: ______________________________

9. Has supported my child in meeting IEP goals and objectives for this year. YES NO NA

COMMENT: ______________________________

Signature of Parent (Optional) ______________ **Date** ______________

Please return to: ______________ (Supervisor) by: ______________ (Date)

Figure 6.2 Continued

1985). Teacher participants in these studies have found the parent data especially helpful for personal and professional growth. In Faucette et al.'s 1995 project, teachers reported that they were not previously aware that parents possessed either the quality or the quantity of information about classroom practices that the surveys confirmed. They generally accepted the parent data as informative, accurate, and useful, and said that they found the comment section of the survey particularly valuable. The parent respondents in this study also reported positive reactions regarding the opportunity to provide the teachers with feedback. The following parent statement is indicative of this finding: "I am proud that my child attends school in a place where educators care about my opinion. I welcome this wonderful opportunity to help my child's teacher."

Figure 6.2 provides another example of a parent survey currently being used by one school district.

Can Students Provide Reliable and Useful Measures of Teaching Effectiveness?

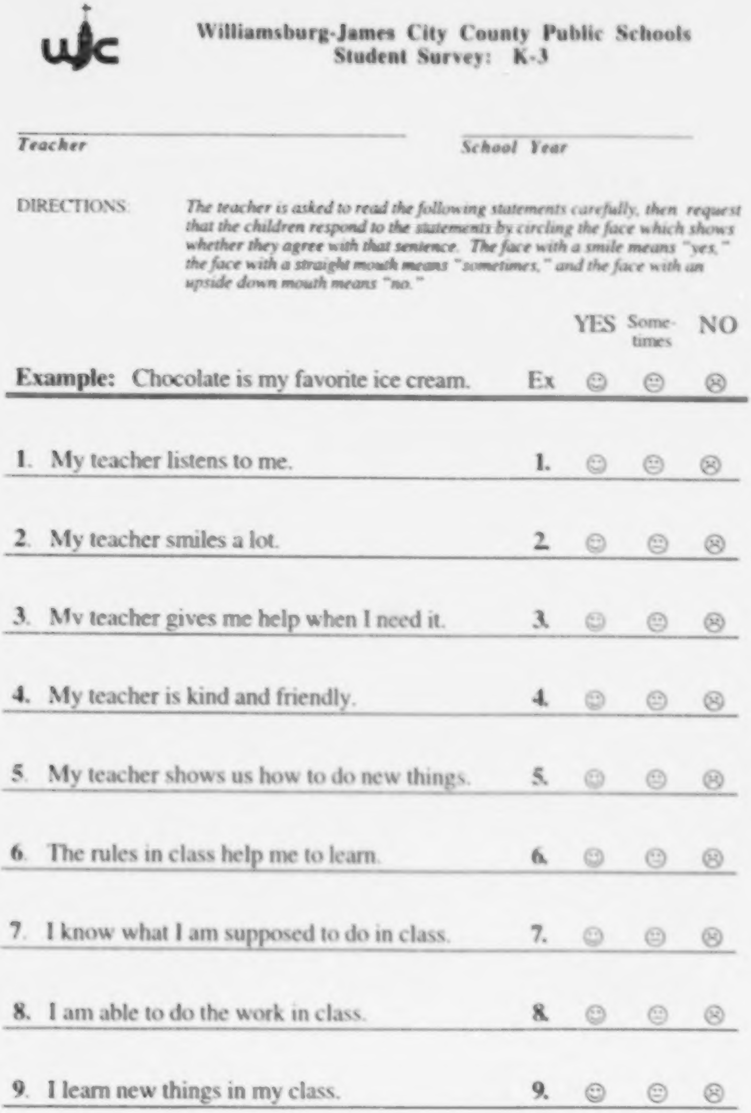
The practice of involving students in teacher evaluation has a long history and appears to be gaining attention (Aleamoni, 1981; Follman, 1995; Harris, 1986). It is one of the most frequently investigated topics in teacher evaluation (Peterson & Kauchak, 1982); however, most of these investigations have been conducted at the college level (Follman, 1996). Many believe that "a focus on pupils reveals far more about the effectiveness of teachers than does direct study of teachers themselves" (McNeil & Popham, 1973, p. 218).

One of the arguments for including students as evaluators is that they are the primary consumers of teachers' services. Students' attitudes toward school and toward classroom practices can contribute to teachers' overall effectiveness or can cause disruption (Northwest Regional Educational Laboratory, 1990). As direct recipients of the teaching-learning process, students are the major clients of teachers, and they are in the key position to provide information about teacher effectiveness. Most important, students are the only ones among all the teacher's clients who have direct knowledge about classroom practices on a regular basis. Student perceptions of quality, therefore, may be more meaningful to the teacher than the judgments of members of any other client group (Peterson & Kauchak, 1982).

Follman (1992) observes that "no other individual or group has [the] breadth, depth, or length of experience with the teacher . . . teachers look to their students rather than to outside sources for indications of their teaching performance" (p. 169). Evidence regarding teacher and principal support for the practice of involving students is mixed, but some see student involvement as beneficial from the student's perspective as well as from the teacher's (Aleamoni, 1981; Follman, 1992). Students can provide information on a teacher's ability to motivate and the degree of communication and rapport that is present. Equally important, student observations of the teacher are unobtrusive (Peterson & Kauchak, 1982). The fact that there is a positive relationship between student ratings and learning is viewed by some as a convincing reason to involve students in the evaluation process (Aleamoni, 1981).

Use of Student Feedback in Teacher Evaluation

Student age and client surveys. As far back as 1925, Flinn (cited in Follman, 1992) found more agreement among the ratings of four groups of high school students than among three groups of principals regarding teacher performance. Follman reviewed more than 20 studies, spanning 70 years, and concluded that secondary students "have and can rate teachers reliably" (p. 171). He also found that variables



Williamsburg-James City County Public Schools
Student Survey: K-3

Teacher ______________________ School Year ______________

DIRECTIONS: *The teacher is asked to read the following statements carefully, then request that the children respond to the statements by circling the face which shows whether they agree with that sentence. The face with a smile means "yes," the face with a straight mouth means "sometimes," and the face with an upside down mouth means "no."*

		YES	Sometimes	NO
Example: Chocolate is my favorite ice cream.	Ex	☺	😐	☹
1. My teacher listens to me.	1.	☺	😐	☹
2. My teacher smiles a lot.	2.	☺	😐	☹
3. My teacher gives me help when I need it.	3.	☺	😐	☹
4. My teacher is kind and friendly.	4.	☺	😐	☹
5. My teacher shows us how to do new things.	5.	☺	😐	☹
6. The rules in class help me to learn.	6.	☺	😐	☹
7. I know what I am supposed to do in class.	7.	☺	😐	☹
8. I am able to do the work in class.	8.	☺	😐	☹
9. I learn new things in my class.	9.	☺	😐	☹

Figure 6.3. Sample of a Student Survey for Grades K-3
SOURCE: Williamsburg-James City County Public Schools. Reprinted by permission.

		YES	Sometimes	NO
10. We have plenty of paper, pencils, and other things to learn with.	**10.**	☺	😐	☹
11. I like school.	**11.**	☺	😐	☹

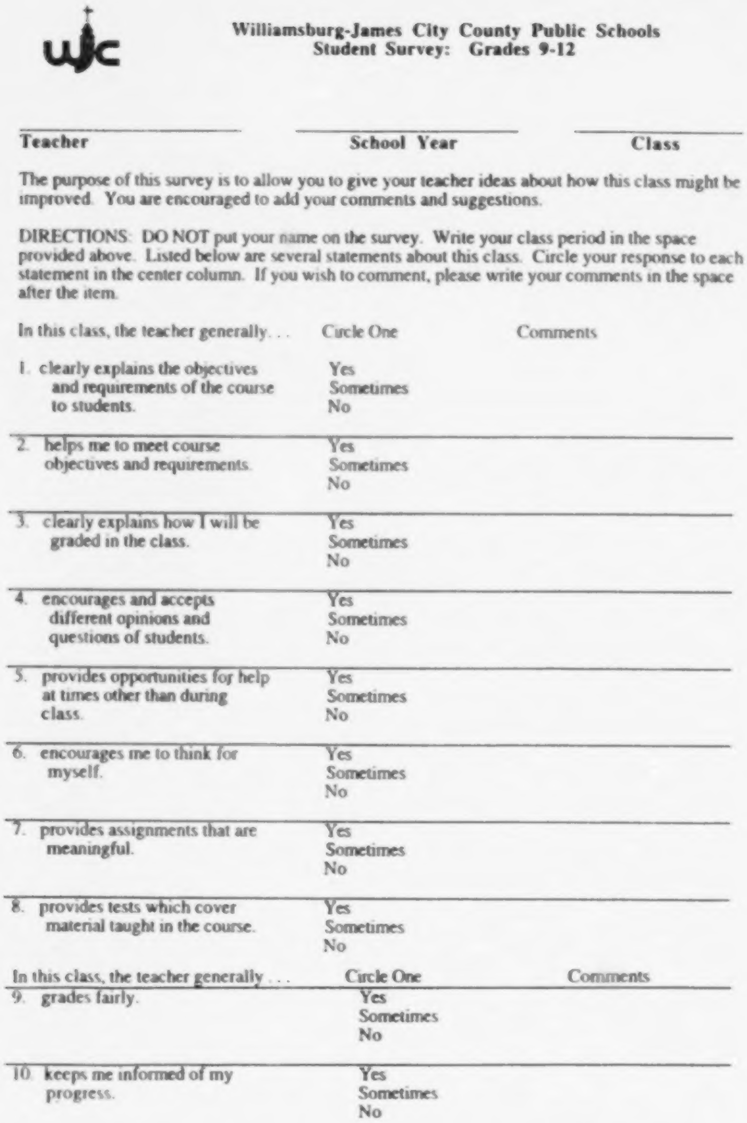
Figure 6.3 Continued

such as age, gender, and ethnicity did not significantly affect the students' ratings. (See Figures 6.3 and 6.4 for examples of rating scales designed for different age groups of students.)

Haak, Kleiber, and Peck (1972) suggest that only students in grades 4 and above should be involved in the evaluation process; however, students as young as kindergarten age have demonstrated adequate reliability for inclusion in the process (Driscoll, Peterson, Browning, & Stevens, 1990). Follman (1992) contends that the issue should ultimately be "whether students have the experience, knowledge, wisdom, judgment, and poise to discriminate and/or evaluate anybody on anything, let alone a professional person such as a teacher, on their performance" (p. 175).

Student ability to assess teacher performance. A frequent perception among teachers is that student ratings are easily influenced and that they are based on popularity, but there is evidence that this concern is unwarranted (Peterson & Kauchak, 1982). Ebmeier, Jenkins, and Crawford (1991) compared high school students' ratings of meritorious and nonmeritorious teachers with ratings from expert practitioners. They conclude that the students were able to discriminate between the two groups as well as were the qualified evaluators.

Despite the potential value of using student feedback in teacher evaluation, some cautions should be noted. McGreal (1994) and Harris (1986) both suggest that student involvement in teacher assessment should be limited to descriptions of life in the classroom and should not include ratings of teacher worth. Peterson and Kauchak (1982) and McNeil and Popham (1973) recommend restrict-



Williamsburg-James City County Public Schools
Student Survey: Grades 9-12

Teacher ____________ **School Year** ____________ **Class** ____________

The purpose of this survey is to allow you to give your teacher ideas about how this class might be improved. You are encouraged to add your comments and suggestions.

DIRECTIONS: DO NOT put your name on the survey. Write your class period in the space provided above. Listed below are several statements about this class. Circle your response to each statement in the center column. If you wish to comment, please write your comments in the space after the item.

In this class, the teacher generally. . .	Circle One	Comments
1. clearly explains the objectives and requirements of the course to students.	Yes Sometimes No	
2. helps me to meet course objectives and requirements.	Yes Sometimes No	
3. clearly explains how I will be graded in the class.	Yes Sometimes No	
4. encourages and accepts different opinions and questions of students.	Yes Sometimes No	
5. provides opportunities for help at times other than during class.	Yes Sometimes No	
6. encourages me to think for myself.	Yes Sometimes No	
7. provides assignments that are meaningful.	Yes Sometimes No	
8. provides tests which cover material taught in the course.	Yes Sometimes No	
In this class, the teacher generally . . .	Circle One	Comments
9. grades fairly.	Yes Sometimes No	
10. keeps me informed of my progress.	Yes Sometimes No	

Figure 6.4. Sample of a Student Survey for Grades 9-12
SOURCE: Williamsburg-James City County Public Schools. Reprinted by permission.

11. sets high standards and expectations for everyone.	Yes Sometimes No
12. encourages students to work together.	Yes Sometimes No
13. uses class time effectively.	Yes Sometimes No
14. manages a classroom that minimizes disruption.	Yes Sometimes No
15. enforces disciplinary rules fairly and consistently.	Yes Sometimes No
16. makes learning interesting and meaningful.	Yes Sometimes No
17. uses a variety of resources, such as technology, when teaching class.	Yes Sometimes No
18. presents material in a variety of ways.	Yes Sometimes No
19. communicates with my parents.	Yes Sometimes No

Please use the space below to comment on the outstanding strengths and/or weaknesses of the teacher and the course. Thank you for your assistance.

Figure 6.4 Continued

ing student data to observations of teachers' discrete and visible behaviors as a way to increase reliability. Although Larson (1984) reports research that "shows that the most valid and reliable opinions regarding teacher performance are solicited from students" (p. 16), he maintains that it should be up to teachers to decide if student ratings should be shared with their supervisors.

BEST COPY AVAILABLE

Purposes of student ratings. Aleamoni (1987) argues that the practice of involving students in teacher evaluation should be restricted to formative evaluation, but the evidence for this restriction is not conclusive. At the very least, Peterson and Stevens (1988) suggest, student data for several years are needed to establish patterns of performance. Even then, these researchers advocate teacher choice concerning the use of student ratings for summative evaluations.

In the 1995 Ostrander study described earlier, the correlation between student and parent ratings was strong. This correlation indicates that collecting data from both groups simultaneously may provide redundant data. Therefore, if both student and parent surveys are to be conducted, it may be better for parents and students to respond to different questions.

How Can Feedback From Peers Be Used Most Effectively in Teacher Evaluation?

Although the use of client input from students and parents can provide important feedback for use in teacher assessment, as with classroom observation this information offers only a partial view of teacher performance. If the intent of the teacher evaluation system is to achieve a full and accurate picture of performance in order to guide improvement and make personnel decisions, then the inclusion of feedback from peers should be considered. Feedback from peers can offer valuable and valid insights into teacher performance.

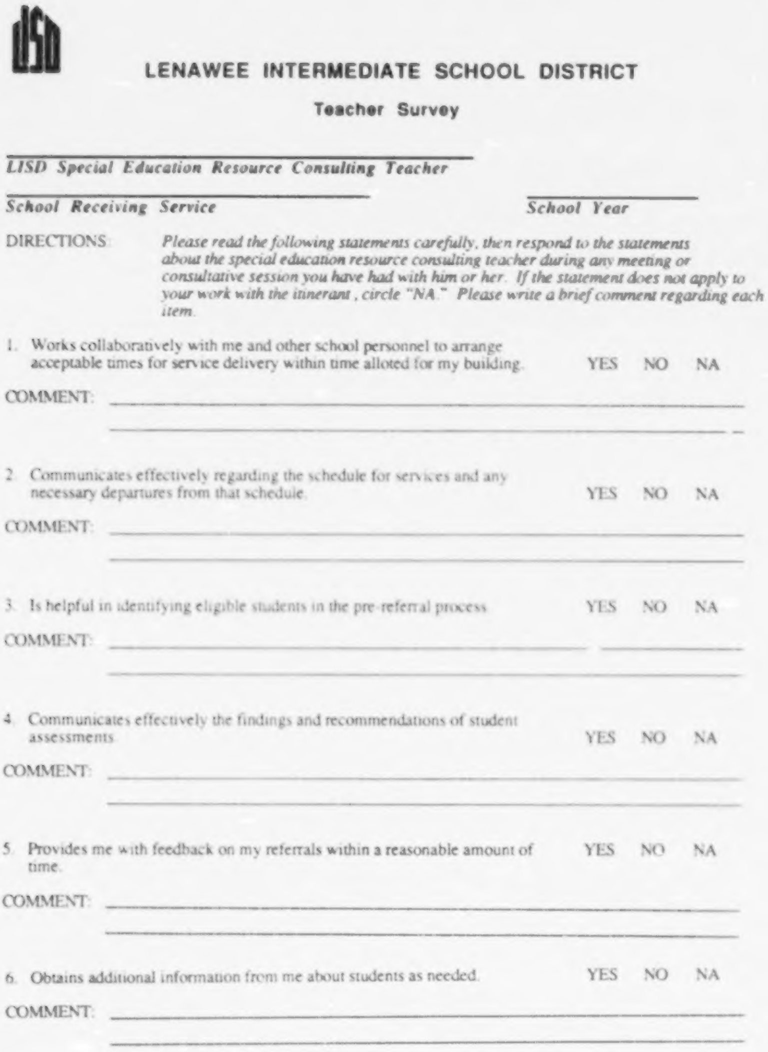
The Case for Peer Feedback

Earlier in this chapter we noted the importance of collaboration in effective teacher evaluation systems—especially if evaluation is used as a tool to enhance quality and support improvement. As a key component of quality improvement efforts, American business and industry have widely adopted the concept of "360 degree" assessment, in which all individuals who interact with an employee (e.g., customers, coworkers, supervisors) provide input into that person's personnel evaluation. The 360 degree concept closely parallels the movement to seek continuous performance improvement by using peer feedback in teacher evaluation.

Teachers often interact with and, in a real sense, serve other teachers and educators in much the same way teachers serve students. Consider, for example, the typical workday of a special education resource consulting teacher. Although the special education teacher may have a caseload for working directly with students, it is likely that a significant portion of his or her day will be spent interacting directly with peers in ways such as coteaching with classroom teachers; conferencing with support staff, such as the school counselor or social worker; and planning with the school principal, teachers, and other staff. This scenario is far from unusual; indeed, collaboration as a primary mode of educational service delivery has come to be the norm rather than the exception for many educators, including special education teachers, gifted/talented resource teachers, reading teachers, and a host of others (e.g., library/media specialists, counselors). Moreover, in recent decades classroom teachers have assumed increasingly collaborative responsibilities and are spending growing amounts of time collaborating directly with other educators.

Given the collaborative nature of the contemporary teacher/educator, it would be unfair and ill-advised to ignore peers as a vital source of performance feedback. Collaborating educators can attest to the quality and quantity of work of the teacher in ways that no one else can. Indeed, "the most complete picture of an employee's performance will be obtained by questioning at least a representative sampling from all of the various constituencies with whom the person works" (Stronge & Helm, 1991, p. 180).

In addition to providing accurate feedback on performance, including peers in a collegial model of assessment can provide professional growth opportunities and encouragement for teachers (see Figure 6.5 for an example of a survey used to gather peer feedback in one school district). As McGreal has noted, "We know that adults respond primarily to positive reinforcement, that they prefer to operate in a collegial and collaborative environment. And traditional teacher evaluation violates many of these new understandings" (quoted in Brandt, 1996, p. 30). Moreover, in a collegial model approach to evaluation "teachers are encouraged to form relationships that enable school change to be based on . . . what is really best for the students" (Haefele, 1992, p. 25).



LENAWEE INTERMEDIATE SCHOOL DISTRICT

Teacher Survey

____
LISD Special Education Resource Consulting Teacher

_______________________________ _______________
School Receiving Service *School Year*

DIRECTIONS: *Please read the following statements carefully, then respond to the statements about the special education resource consulting teacher during any meeting or consultative session you have had with him or her. If the statement does not apply to your work with the itinerant, circle "NA." Please write a brief comment regarding each item.*

1. Works collaboratively with me and other school personnel to arrange acceptable times for service delivery within time alloted for my building. YES NO NA

COMMENT: _______________________________________

____

2. Communicates effectively regarding the schedule for services and any necessary departures from that schedule. YES NO NA

COMMENT: _______________________________________

____

3. Is helpful in identifying eligible students in the pre-referral process. YES NO NA

COMMENT: _______________________________________

____

4. Communicates effectively the findings and recommendations of student assessments. YES NO NA

COMMENT: _______________________________________

____

5. Provides me with feedback on my referrals within a reasonable amount of time. YES NO NA

COMMENT: _______________________________________

____

6. Obtains additional information from me about students as needed. YES NO NA

COMMENT: _______________________________________

____

Figure 6.5. Sample of a Peer Survey
SOURCE: Lenawee Intermediate School District. Reprinted by permission.

7. Works cooperatively with me and other school personnel to identify and achieve student IEPC goals and objectives. YES NO NA

COMMENT: ______________________________

8. Communicates effectively with me regarding student progress. YES NO NA

COMMENT: ______________________________

9. Enlists the support of other school personnel, parents, and community resources when necessary to meet the needs of students. YES NO NA

COMMENT: ______________________________

10. Demonstrates appropriate professional conduct. YES NO NA

COMMENT: ______________________________

11. Is considerate of my time. YES NO NA

COMMENT: ______________________________

12. Makes me feel comfortable about asking for help in the future. YES NO NA

COMMENT: ______________________________

Signature of Teacher (Optional) **Date**

Please return to: ______________ (Supervisor) by: ______________ (Date)

Figure 6.5 Continued

A Caution About Using Peer Feedback

Although soliciting peer feedback is a promising practice for use in evaluating teachers, there are pitfalls that must be avoided if peer information is to be meaningful. To begin, evaluators should recognize that the information gained by inviting peers (or, for that matter, students or parents) to provide feedback regarding teacher performance is merely that—feedback. Peer feedback is not the same as peer evaluation. Information generated by peers through a client survey or another data collection format is only one data source to be used as part of a comprehensive teacher evaluation system.

BEST COPY AVAILABLE

Peterson (1995) identifies two additional concerns with peer-generated information in teacher evaluation. First, he notes that "a superficial survey of fellow teachers has serious conflict-of-interest problems"; second, "a serious drawback to peer surveys is the lack of direct knowledge that teachers have of each other's practice" (p. 175). If soliciting peer feedback will result only in a measure of perception and will shed no light on direct performance of the teacher in fulfilling his or her job, then there is no value added to the evaluation process. Failure to account for and minimize (if not eliminate) these concerns could not only result in the gathering of inappropriate/invalid information, it could be detrimental to the total teacher evaluation effort.

Guidelines for Using Peer Feedback

If peers are to be invited into the data collection process for teacher evaluation, several issues should be considered if the above-named concerns are to be overcome and the value of peer feedback maximized. In particular, the issues of factual data use, validity and reliability considerations, confidentiality, and sampling procedures should be addressed. Each is explored briefly below.

Use of factual data. In eliciting evaluation data from peers and other clients, "the single most important consideration is to focus on obtaining factual descriptions of the person's behavior or performance" (Stronge & Helm, 1991, p. 193). One important way to avoid receiving perceptual information is to invite feedback only from peers who work directly with or have direct knowledge of a given teacher's performance. Another consideration in generating factual data is to ask only for information that the teacher's peers are in a position to provide. "Every attempt . . . should be made to develop survey instruments or to ask questions that focus on specific, desired activities or behaviors related to the . . . employee's job" (p. 194).

Validity and reliability considerations. A client survey is valid only to the extent that it actually measures what it purports to measure. For peer feedback, one simple way to increase validity is to ask only job-related questions that the respondents are in a position to answer.

For example, evaluators should not ask a peer-teacher how well he or she feels the evaluated teacher-colleague is performing in the area of classroom management if the peer has no direct knowledge of that performance.

Reliability, the degree to which a measure produces consistent results, is another issue that deserves attention. Generally, the larger the sample, the higher the reliability. One of the major drawbacks of formal classroom observation is the extremely small sample size; with only one or two observations, the results of the observations would be suspect. Peer feedback typically avoids this problem because multiple responses are provided in a client survey. However, the reliability of client surveys also is affected by the extent of past use. The results of one or two sets of surveys may be suspect, but as client surveys are administered repeatedly over time, more confidence typically can be placed in the findings. Thus, evaluators should be cautious about placing too much trust in the results of a single data set.

Confidentiality. Peers should be guaranteed anonymity and the teacher and peers should be assured of confidentiality in the collection and use of peer feedback. Strict confidentiality should be maintained throughout the teacher evaluation process. Otherwise, "few individuals will respond with candor—especially subordinates or peers who might be concerned about jeopardizing their own working relationship with the employee" (Stronge & Helm, 1991, p. 181).

Sampling procedures. A client survey is virtually useless without a good sample. One cannot assume that the findings are indicative of performance without reasonably good sampling procedures. With inappropriate sampling procedures, a problem similar to the one that Goldilocks encountered in her visit to the three bears would occur: Would the inferences drawn from the sample be too positive, too negative, or just about right? This concern, in fact, holds true for classroom observations, analysis of artifacts of performance, and any other data source. Thus, if any type of client survey is to be conducted, the evaluator must give careful attention to which individuals will be surveyed and how that information will be collected.

Using Client Surveys as Part of Multiple Data Sources in Teacher Evaluation

Contemporary research on teacher evaluation has emphasized the importance of the development of multiple and variable lines of evidence of teacher performance to improve the state of teacher evaluation (Darling-Hammond, Wise, & Pease, 1983; Harris, 1986; Peterson, 1995; Scriven, 1994; Stanley & Popham, 1988; Stronge, 1995; Stronge & Helm, 1991). This movement parallels the national trend toward increased client involvement in school governance and decision making.

The knowledge base regarding effective multiple lines of evidence is, however, inadequate, and many potential data sources are relatively untested (Epstein, 1985; Peterson, 1995; Peterson, Gunne, Miller, & Rivera, 1984; Scriven, 1981). Abrami and d'Apollonia (1990) have approached the question of the validity of client ratings from two perspectives: If assessments correctly reflect clients' opinions about service quality, then their views are worth knowing and thereby provide a valid measure of satisfaction; on the other hand, the same clients' opinions may not provide a valid measure of the effectiveness of the service. For example, parent ratings may provide a teacher with valid information about client satisfaction with a course but not with meaningful information about teaching quality. Abrami and d'Apollonia recommend correlating judgments from multiple sources to establish validity for client information.

Ostrander's (1995) study, discussed earlier in this chapter, provides evidence that the use of multiple judges may provide unique perspectives on teacher performance. The findings of the study are consistent with Epstein's (1985) conclusion: "Because there is not a single set of skills that perfectly define effective teaching, measures of many aspects of teaching by multiple judges are likely to yield the fairest and most comprehensive evaluation of teachers" (p. 8).

It is unrealistic to expect any one person to be fully cognizant of how effectively all subordinates carry out assigned duties and responsibilities. Using multiple data sources allows administrators to focus more efficiently the time they invest in evaluation and to cultivate more accurate decisions about teacher competency (Tucker, Bray, & Howard, 1989).

Summary

Scriven (1994) contends that to be a good teacher, one must have a commitment to success and to improvement. You may accomplish this by "using serious self-evaluation and, for objectivity, subjecting yourself to evaluation by others chosen or accepted by you." Scriven adds, "accountability obliges you to be able to demonstrate your success to third parties—not merely to your own satisfaction" (p. 159).

Although additional empirical evidence is needed to determine how multiple measures of performance received from multiple sources contribute to the process of teacher evaluation (Epstein, 1985; McGreal, 1994; Ostrander, 1995; Peterson, 1987), it is clear that the use of a multifaceted data collection process can improve overall data quality. Additionally, the use of client surveys as part of a comprehensive teacher evaluation system can provide administrators and teachers with better feedback and assessment information, both for personal and professional improvement and for ensuring accountability in performance.

Note

1. The grand mean is an overall average of all item means calculated out of a possible 4.0.

References

Abrami, P. C., & d'Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. In M. Theall & J. Franklin (Eds.), *Student ratings of instruction: Issues for improving practice* (pp. 97-103). San Francisco: Jossey-Bass.

Aleamoni, L. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 110-145). Beverly Hills, CA: Sage.

Aleamoni, L. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education, 1,* 111-119.

Bayless, D. L., Massaro, G., Bailey, E., Coley, D., Holladay, R., & McDonald, D. (1992). The quality improvement management approach as implemented in a middle school. *Journal of Educational Psychology, 6,* 191-209.

Becher, R. M. (1984). *Parent involvement: A review of research and principles of successful practice* (Report No. PS 014 563). Urbana, IL: National Institute of Education. (ERIC Document Reproduction Service No. ED 247 032)

Bonstingl, J. J. (1992). *Schools of quality: An introduction to total quality management in education*. Alexandria, VA: Association for Supervision and Curriculum Development.

Brandt, R. (1996). On a new direction for teacher evaluation: A conversation with Tom McGreal. *Educational Leadership, 53*(6), 30-33.

Bridges, E. (1986). *The incompetent teacher*. Philadelphia: Falmer.

Costa, A. (Speaker). (1993). *Leadership: Protecting the intellectual ecology of the school* (Cassette Recording No. 93-4134). Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53*, 285-328.

Deming, W. E. (1986). *Out of the crisis*. Cambridge: MIT Press.

Dixon, R. G. D. (1994). Future schools: And how to get there from here. *Phi Delta Kappan, 75*, 360-365.

Driscoll, A., Peterson, K., Browning, M., & Stevens, D. (1990). Teacher evaluation in early childhood education: What information can young children provide? *Child Study Journal, 20*, 67-69.

Ebmeier, H., Jenkins, R., & Crawford, G. (1991). The predictive validity of student evaluations in the identification of meritorious teachers. *Journal of Personnel Evaluation in Education, 4*, 341-347.

Educational Research Service. (1988). *Teacher evaluation: Practices and procedures*. Arlington, VA: Author.

Ellett, C. D., & Garland, J. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to "state-of-the-art" systems? *Journal of Personnel Evaluation in Education, 1*, 69-92.

Epstein, J. (1985). A question of merit: Principals' and parents' evaluation of teachers. *Educational Researcher, 14*(7), 3-10.

Evertson, C. M., & Holley, F. M. (1981). Classroom observation. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 90-109). Beverly Hills, CA: Sage.

Faucette, S. L., Ball, D. M., & Ostrander, L. R. (1994). Parents as valued customers: The Virginia Beach Parent Perception Survey. *ERS Spectrum, 12*(2), 3-9.

Faucette, S. L., Ball, D. M., & Ostrander, L. R. (1995). *Client satisfaction surveys* (Unpublished executive summary). Virginia Beach, VA: Virginia Beach City Public Schools.

Follman, J. (1992). Secondary school students' ratings of teacher effectiveness. *High School Journal, 75,* 168-178.

Follman, J. (1995). Elementary public school pupil rating of teacher effectiveness. *Child Study Journal, 25,* 57-78.

Follman, J. (1996). *Pupil ratings of teacher effectiveness.* Paper presented at the CREATE National Evaluation Institute, Bethesda, MD.

Frase, L., & Streshly, W. (1994). Lack of accuracy, feedback, and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education, 1,* 47-57.

Fullan, M. (Speaker). (1994). *Harnessing the forces of educational reform* (Cassette Recording No. 94-GS03). Alexandria, VA: Association for Supervision and Curriculum Development.

Goodlad, J. I. (1984). *A place called school.* New York: McGraw-Hill.

Haak, R. A., Kleiber, D., & Peck, R. (1972). *Student evaluation of teacher instrument, II.* Austin: Texas University, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 080 574)

Haefele, D. (1992). Evaluating teachers: An alternative model. *Journal of Evaluation in Education, 5,* 335-345.

Haefele, D. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education, 7,* 21-31.

Harris, B. (1986). *Developmental teacher evaluation.* Boston: Allyn & Bacon.

Larson, R. (1984). Teacher performance evaluation: What are the key elements? *NASSP Bulletin, 68*(469), 13-18.

Mager, G. M. (1980). The conditions which influence teachers in initiating contacts with parents. *Journal of Educational Research, 73,* 276-282.

Mark, M. M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review, 9,* 605-626.

McGreal, T. L. (Speaker). (1994). *The next generation of teacher evaluation* (Cassette Recording No. 94-4625). Alexandria, VA: Association for Supervision and Curriculum Development.

McNeil, J. D., & Popham, W. J. (1973). The assessment of teacher competence. In R. M. W. Travers (Ed.), *Second handbook of research in teaching* (pp. 218-243). Chicago: Rand McNally.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.

Northwest Regional Educational Laboratory. (1990). *Effective schooling practices: A research synthesis*. Portland, OR: Northwest Regional Educational Laboratory.

Ostrander, L. P. (1995). *Multiple judges of teacher effectiveness: Comparing teacher self-assessments with the perceptions of principals, students, and parents*. Unpublished doctoral dissertation, University of Virginia, Charlottesville.

Peterson, K. D. (1987). Teacher evaluation with multiple and variable lines of evidence. *American Educational Research Journal, 24*, 311-317.

Peterson, K. D. (1988). Parent surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education, 2*, 239-249.

Peterson, K. D. (1995). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin.

Peterson, K. D., Gunne, G., Miller, P., & Rivera, O. (1984). Multiple audience rating form strategies for student evaluation of college teaching. *Research in Higher Education, 20*, 309-321.

Peterson, K. D., & Kauchak, D. (1982). *Teacher evaluation: Perspectives, practices, and promises* (Report No. SP 022 900). Salt Lake City, UT: Center for Educational Practice. (ERIC Document Reproduction Service No. ED 233 996)

Peterson, K. D., & Mitchell, A. (1985). Teacher-controlled evaluation in a career ladder program. *Educational Leadership, 43*(3), 44-47.

Peterson, K. D., & Stevens, D. (1988). Student reports for school teacher evaluation. *Journal of Personnel Evaluation in Education, 2*, 19-31.

Root, D., & Overly, D. (1990). Successful teacher evaluation: Key elements for success. *NASSP Bulletin, 74*(527), 34-38.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 244-271). Beverly Hills, CA: Sage.

Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8*, 151-184.

Stanley, S. J., & Popham, W. J. (Eds.). (1988). *Teacher evaluation: Six prescriptions for success*. Alexandria, VA: Association for Supervision and Curriculum Development.

Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation, 21*, 131-151.

Stronge, J. H., & Helm, V. M. (1991). *Evaluating professional support personnel in education*. Newbury Park, CA: Sage.

Stronge, J. H., Helm, V. M., & Tucker, P. D. (1995). *Evaluation handbook for professional support personnel*. Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Tetenbaum, T., & Mulkeen, T. (1988). Assessment of educational personnel in the twenty-first century. *Journal of Personnel Evaluation in Education, 1*, 235-244.

Tucker, N. A., Bray, S. W., & Howard, K. C. (1989). Using a client-centered approach in the principal's evaluation of counselors. *Journal of Personnel Evaluation in Education, 2*, 335-353.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices*. Santa Monica, CA: RAND Corporation.

Wood, C. (1992). Toward more effective teacher evaluation: Lessons from naturalistic inquiry. *NASSP Bulletin, 76*(542), 52-59.

7

Indicators of Student Learning in Teacher Evaluation

Andrew A. McConney
Mark D. Schalock
H. Del Schalock

The purpose of teaching is learning, and the purpose of schooling is to ensure that each new generation of students accumulates the knowledge and skills needed to meet the social, political, and economic demands of adulthood. Yet few regular evaluation systems for teachers take student learning into account, and few regular evaluation systems for schools or school programs take more than student progress in basic knowledge and skills into account—usually as these are measured through state-developed or commercially prepared standardized tests of achievement.[1] Although the idea of linking pupil learning to teacher and school evaluation/improvement has had strong appeal and numerous adherents over the years (Haertel, 1986; Millman, 1981; Millman & Sykes, 1992; Scriven, 1988; Smylie, 1994; Webster, 1994), the technical and political problems encountered in attempting to make these linkages in ways that are meaningful and acceptable to all concerned have prevented their widespread adoption.

In this chapter, our purpose is to outline two emerging and promising approaches to using indicators of student learning in the evaluation of teachers' work. We also provide context for our conclusions and recommendations by describing (a) past uses to which measures of student progress in learning have been put in evaluating the performance of teachers and (b) current approaches to measuring student learning and the challenges and potential each of these holds for the evaluation of teachers. We close the chapter by offering what we view as a balanced, sensible approach to the inclusion of student achievement data in systems for teacher (and school) evaluation. As part of this discussion we briefly consider our understanding of the broad principles that seem to be gaining acceptance as guides to the design of teacher evaluation systems that include evidence of student progress in learning.

Past Efforts and Their Consequences

Efforts in the United States during the past 50 years to connect pupil learning meaningfully to teacher work have followed many paths.[2] Most of these have been pursued by researchers concerned with teacher or school effects. Others, however, have been pursued by practitioners or policy makers concerned with enhancing the productivity of teachers and schools. Table 7.1 provides a list of six major efforts that have focused on one or the other of these goals as well as synopses of their consequences. Though anchored to pupil learning, none of the efforts summarized in Table 7.1 has made a dramatic difference in the effectiveness of schools, and all have helped shape a legacy that has led both researchers and practitioners to doubt whether school-based learning can ever be connected meaningfully and defensibly to appraising the work of teachers.

Two features of these various research and development efforts have combined to limit their contribution to school improvement and to feed the suspicion of skeptics about the defensibility of linking the learning of students to evaluating the work of teachers. One is the heavy misuse of standardized achievement tests to measure student progress in learning, even though the limitations of these measures as indicators of school-based learning have been known for decades. As noted by Sanders and Horn (1995), "The misuse and under-utilization of [standardized] test data in the past haunts the

TABLE 7.1 Six Lines of Research Around Influences on Student Learning and Their Consequences

Focus	*Consequence*
Research on teacher effects, particularly process-product research	Gave credence to the idea that student performance on *district-administered* standardized achievement tests taken at the end of a school year can be looked to as defensible measures of teacher effectiveness and, in doing so, helped galvanize teachers' criticism of such measures
Research on school effects	Gave credence to the idea that improving a school as a place for learning is a reasonably straightforward task, and that it can be done by attending to factors other than the competence of teachers
Research on time, class size, and resources for learning	Weakened the policy and political arguments for funds to reduce class size, lengthen the school day or year, or add resources for instruction
Teacher evaluation for merit pay or career ladder advancement	Added to teacher skepticism, and growing criticism, of attempts to link learning gains to teacher work
School evaluation for accountability and organizational control	Added to teacher skepticism, and growing criticism, of attempts to link learning gains to teacher work
Teacher and school evaluation as a central feature of performance contracting	Added to teacher skepticism, and growing criticism, of attempts to link learning gains to teacher work

present." A second feature of such efforts is that most have failed to take into account the conditions under which teaching and learning occur, and often the characteristics of the students being taught.

These important factors have combined to make such measures of student progress in learning distant from the day-to-day work of teachers and insensitive to the conditions under which teaching and learning occur. As a consequence, they are seen as having little meaning for or value to teachers, and only modest value to students and their parents (beyond the value that large-scale assessments have for comparing a child or group of children with others). Under these conditions, it should come as no surprise that teachers are skeptical about, and often openly hostile toward, the use of externally developed measures of student learning to assess their effectiveness as professionals. As new state assessment systems are developed and implemented, these circumstances may change, but great damage already has been done. Many researchers and school professionals now hold the view—with some justification, given past efforts—that attempting to connect learning gains by students to the work of individual teachers for purposes of assessment or evaluation is indefensible.

In addition to being suspect for methodological reasons, previous attempts to connect student learning to the appraisal of teacher work have carried with them features of schooling that are inconsistent with many of the current emphases teachers are encountering in restructured schools. For most of the past century, these features have included the following:

- Teaching and learning dominated by textbooks
- Learning outcomes centered on factual knowledge or basic skills
- Pupil learning assessed through multiple-choice, true/false, or other short-answer paper-and-pencil tests
- Performance on teacher-administered examinations used primarily for purposes of grading rather than the evaluation of instruction or improvement of learning

Increasingly, in today's schools, the following changes have taken or are taking place:

- Teachers view textbooks as resources for learning rather than definers of what is to be learned.

- Outcomes of importance for students include reasoning, problem solving, and other forms of knowledge and skill *application* as well as their acquisition.
- "Authentic" means of assessment are emphasized whenever possible and appropriate, and are merged with more traditional approaches to assessment in portfolios of student work.
- Curriculum, instruction, and assessment are viewed as *interdependent and aligned* components of effective teaching rather than as components to be thought about separately, and often independently.

Any teacher evaluation system used in today's schools that attempts to include in it evidence of student progress in learning must be designed to take these fundamental changes in the nature of student and teacher work into account.

Which Emerging Systems Are Promising?

The skepticism, suspicion, and outright rejection growing from past efforts to include student learning in the appraisal of teachers' work has not prevented new efforts from going forward. The logic of the connection, and the potential for school improvement it holds, is simply too strong for educators to lay it aside. Also, new models of schooling, new approaches to the assessment of student learning, and new capacities for managing data pertaining to student learning provide new ways of thinking about the issue.

Perhaps of even greater importance is the emergence of new ways of thinking about teaching and learning as well as new ways of thinking about teacher effectiveness and school productivity. These bring with them a fundamental shift in how we think about education generally, and in the nature of teaching and teacher effectiveness specifically. Space constraints do not permit a discussion of these evolving views in any detail, but they will become reasonably apparent as we describe in the next several pages evolving approaches for including pupil learning as one source of evidence in the evaluation of teachers' performance.

The two approaches described below are classified according to the taxonomy portrayed in Figure 7.1, and are represented toward one end of the continua by systems such as the Dallas (Texas) Public

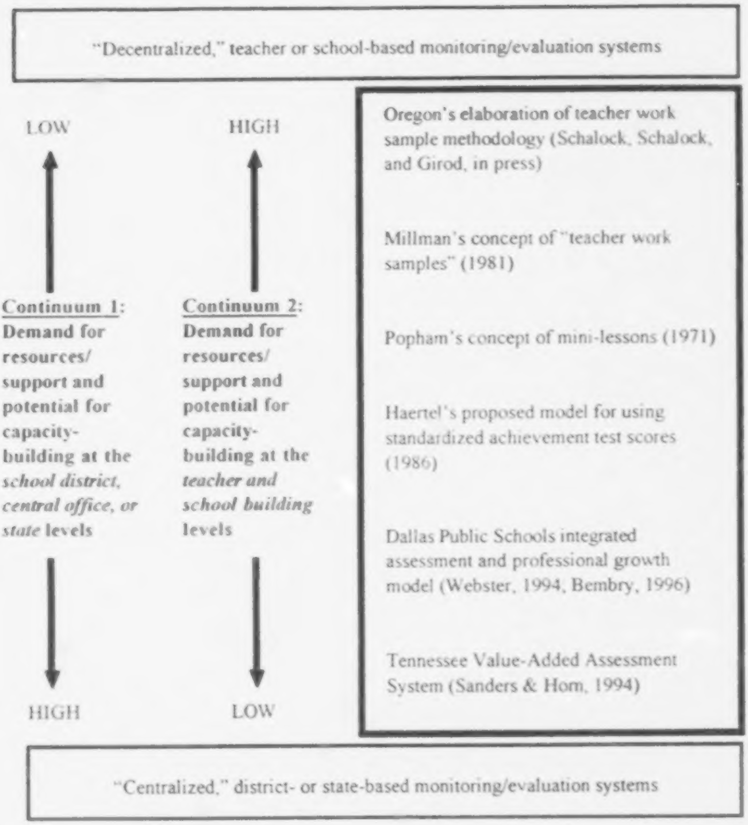

"Decentralized," teacher or school-based monitoring/evaluation systems
LOW
HIGH
Continuum 1: Demand for resources/ support and potential for capacity-building at the *school district, central office, or state* levels
Continuum 2: Demand for resources/ support and potential for capacity-building at the *teacher and school building* levels
HIGH
LOW
Oregon's elaboration of teacher work sample methodology (Schalock, Schalock, and Girod, in press)
Millman's concept of "teacher work samples" (1981)
Popham's concept of mini-lessons (1971)
Haertel's proposed model for using standardized achievement test scores (1986)
Dallas Public Schools integrated assessment and professional growth model (Webster, 1994, Bembry, 1996)
Tennessee Value-Added Assessment System (Sanders & Horn, 1994)
"Centralized," district- or state-based monitoring/evaluation systems

Figure 7.1. A Taxonomy for Differentiating Among Efforts to Apply Student Learning in the Evaluation of Teacher Work

Schools Teacher Evaluation System (Teacher Evaluation Task Force, 1995, 1996; Webster, 1995) and the Tennessee Value-Added Assessment System (TVAAS; Sanders & Horn, 1992, 1994) and at the other end by Oregon's teacher work sample methodology (e.g., McConney & Schalock, 1996; Schalock, Schalock, & Girod, in press; Schalock, Schalock, Myton, & Girod, 1993).[3] The Dallas and Tennessee systems are grouped at one end of the continua in Figure 7.1 because, in our view, these approaches exemplify educational monitoring and evaluation in which the demand for support/resources and the potential for evaluation capacity building are tilted toward a central administration (in Dallas at the district level and in Tennessee at the state level). On the other hand, we have placed Oregon's teacher work sample methodology at the opposite end of the continua because we believe this system exemplifies approaches in which resource demand and capacity-building payoff are tilted toward the individual school and centered on teachers.

Centralized Monitoring and Evaluation Systems

The foundation of both the Dallas Public Schools and Tennessee evaluation systems is a value-added methodology that allows the computation of "effectiveness indices."[4] These indices essentially reflect the performance and accomplishments of school systems, schools, and teachers in terms of students' growth in learning over time (Sanders & Horn, 1992, 1994; Teacher Evaluation Task Force, 1995, 1996; Webster, 1995; Webster & Mendro, 1995; Webster, Mendro, Bembry, Orsak, & Weerasinghe, 1995). In both systems, student learning is measured by a series of large-scale, national- or state-normed standardized tests given at the end of each school year, generally in April and May (in Dallas Public Schools these include Assessments of Course Performance [ACP], Iowa Test of Basic Skills [ITBS], Spanish Assessment of Basic Education [SABE], and Texas Assessment of Academic Skills [TAAS]; in Tennessee, the Tennessee Comprehensive Assessment Program [TCAP]).

Dallas has designed a system based on multiple linear regression and hierarchical linear modeling, statistical procedures that translate students' year-to-year performance on the assessments listed into indices that reflect the unique effects of schools and teachers on growth in learning. The Dallas system also addresses, and effectively solves, one of the reasons for the failure of past efforts to employ student learning effectively in the appraisal of teaching, adjusting students' performance scores to take into account student and school contextual factors beyond the control of the teacher or school. Such variables include student ethnicity/language proficiency; student gender; student incoming ability; student socioeconomic status; school overcrowdedness and mobility; school-level average family income; school-level average poverty index; school-level percentage of students receiving free/reduced-price lunches; school-level percentage of limited-English-proficient students; school-level percentages of African American, Hispanic, and other minority students; and school-level percentage of instructional days lost to medical disability leave or unfilled vacancies. In Dallas, this set of background variables (not including student incoming ability) typically accounts for between 15% and 20% of the variance seen in student learning (Webster, 1994, 1995).

In Tennessee, the TVAAS utilizes a mixed-model statistical methodology for assessing the impact of educational systems, schools, and teachers on students' learning growth from year to year using norm-referenced achievement tests (Sanders & Horn, 1992, 1994). Different from the Dallas system, however, the TVAAS bases calculation of system, school, and teacher effects on a longitudinal data set; each student is tested annually in grades 2 through 8 (subject-area assessments are under development for the high school grades). Reports of effects at any level are, therefore, based on at least 3 years of data and no more than 5 years of data (Sanders & Horn, 1992). The TVAAS thus constructs an individual learning profile for each child in the Tennessee system that in essence *includes* the unique background or contextual information associated with that child. In this way, by aggregating data on students over several years and by employing mixed-model statistical methodology, the TVAAS provides unbiased measures of the influence of school systems, schools, and teachers on students' academic progress without the inclusion of an additional set of contextual variables (Sanders & Horn, 1992, 1994).

It is important to recognize that, in Dallas, teachers' effectiveness indices (currently called Classroom Effectiveness Indices) are designed and intended to identify teacher improvement needs and to inform individual teacher improvement plans that are aligned with the school's improvement objectives. This is possible because each teacher's effectiveness index is reported to the school principal and teacher on a student-by-student, class-by-class basis, thus providing both teacher and administrator the data they need to identify the teacher's strengths and areas that need improvement. Effectiveness indices feed into improvement plans, the expectation being that the "more needs identified or the more severe the needs identified, the more extensive will be the specification of remedies for the needs" (Teacher Evaluation Task Force, 1996, p. 8).

From the perspective of the continua depicted in Figure 7.1, both the Dallas Public Schools teacher evaluation system and the teacher appraisal applications of the TVAAS are educational monitoring/ evaluation systems that place heavy data collection, software development, statistical analysis, and reporting demands at the state, district, and/or central office levels. Indeed, Sanders and Horn (1992) have characterized the TVAAS as an "enormous computing task" (p. 2). Not surprisingly, then, the payoff in developing educational monitoring expertise and evaluation capacity comes also largely at the district or central office level, as it is here that data are collected, software developed, statistical analyses conducted, and system, school, and teacher effects reported. Within such an arrangement, both teachers and school administrations are cast primarily in the roles of data receivers and interpreters, as opposed to data providers and analyzers, with little capacity being gained as professionals in collecting and using data on student learning to improve practice. An unintended consequence of this approach to managing information about student learning is that it invites the public to place greater trust in state or central office personnel than in teachers and principals to ensure the effective performance of schools and teachers.

Decentralized, Teacher- or School-Based Monitoring and Evaluation Systems

Oregon's teacher work sample methodology (TWSM) has been developed as a method for the evaluation of prospective teachers for initial licensure and of practicing (early career) teachers for continued licensure.[5] The application of TWSM to the performance appraisal of practicing teachers currently is only under design and study, but we see a number of reasons to provide a description of the methodology here, all of which are associated with the above-noted challenges of applying student learning data to teacher appraisal.

First, unlike the approaches taken in Dallas and Tennessee, TWSM does not require a centralized or sophisticated statistical system for its analysis. Second, the methodology is designed to be a "close to a teacher's work" method (Schalock, Schalock, & Girod, in press), thus avoiding the oft-heard criticism that teacher evaluation systems that rely on large-scale standardized assessments are in danger of creating a mismatch between learning outcomes (goals and objectives) of local importance and value and more global learning outcomes associated with large-scale assessments (see also Tables 7.2 and 7.3, later in this chapter).[6]

Simply put, TWSM requires teachers to think about, develop, implement, document, and present samples of their work as evidence of their effectiveness. As used in Oregon, a central and essential feature of teacher work samples is the inclusion of evidence of stu- dent progress, or lack thereof, in learning. In this authentic, applied performance appraisal procedure, teachers are asked to take the following steps:

1. Define the sample of teaching and learning to be described.
2. Identify the learning outcomes to be accomplished within the work to be sampled.
3. Assess the status of students prior to instruction with respect to the postinstruction outcomes to be accomplished.
4. Align instruction and assessment, through the development of instruction and assessment plans, with learning outcomes to be accomplished and the current status of students with respect to outcome attainment.
5. Describe the context in which teaching and learning are to occur.
6. Adapt outcomes desired and related plans for instruction and assessment to accommodate the demands of the teaching-learning context.
7. Implement a developmentally and contextually appropriate instructional plan.
8. *Assess the postinstructional accomplishments of learners and calculate on a student-by-student basis the growth in learning achieved.*
9. Summarize, interpret, and reflect on student growth in learning in light of the preinstructional developmental levels of students and the context in which teaching and learning occurred.

We have purposely italicized Step 8 to emphasize that TWSM makes explicit to all constituencies the inclusion and prominence of students' learning gains in the definition of "teacher effectiveness." Also, as seen in Steps 5, 7, and 9, the method allows and encourages teachers to take into account the multiple contexts (student, classroom, school, and district) of the teaching-learning situation. TWSM is thus a complex, authentic, appiied performance approach to the appraisal of teachers.

When complete, a teacher's work sample can be viewed as a compact, delimited teacher portfolio, with some important differences from portfolios as generally conceived (Wheeler, 1994; Wolf, 1994). For example, a portfoiio typically represents a fairly long span of time (e.g., an entire school year, with supporting materials from previous years), whereas a teacher work sample brings a fine focus to a shorter period of teaching and learning. More important, whereas a teacher portfolio can include a broad representation of a teacher's work and professional development, TWSM is designed to focus teachers on the following issues and their alignment:

- What are the learning outcomes I want my pupils to accomplish?
- What activities and instructional methodologies are appropriate/necessary for *these* pupils to achieve *these* outcomes?
- What resources and how much time do I need to implement these activities/methodologies?
- What assessment activities/methodologies are appropriate for these pupils and these outcomes when using these instructional methodologies?
- How successful was I at helping my students achieve the learning outcomes desired?
- What went right? What went wrong? Why? (Teaching Research Division, 1995)

By employing this system, teachers and the evaluators of teachers are compelled to examine the context of teaching-learning, what is planned, what is taught, what is assessed, *and what is learned*. Further, the teacher is encouraged to interpret and reflect on her or his success in fostering students' accomplishment of the targeted learning goals. The appraisal of these various work sample components provides significant formative feedback to the teaching professional that allows focused improvement along the dimension(s) indicated.

There is as yet insufficient evidence with which to judge the summative applications of TWSM as part of an evaluation system for making high-stakes personnel decisions. We emphasize, however, that TWSM was never designed or intended as a stand-alone performance evaluation system, and work samples constitute only one of many sources of evidence used in making teacher licensure decisions in Oregon.

As depicted in Figure 7.1, "close to a teacher's work" assessment and evaluation systems such as TWSM place the demands for data collection, analysis, interpretation, reflection, and reporting on the educational system at the teacher and school building levels. There obviously are costs involved in such a monitoring/evaluation system, including initial costs associated with instructing and preparing teachers for the development and production of samples of their work for evaluation. Also, as Wheeler (1994) has noted in the case of teacher portfolios, there are the additional costs that go along with the effort and care needed to train those evaluating teacher work samples (administrators and/or panels of "expert" teachers) to ensure that the scoring of these work samples is accurate and consistent. These are essentially *teacher capacity-building costs*, however, and should be viewed differently from centralized district or state costs involved in managing and reporting large-scale assessment programs.

Similarly, the payoff from evaluation systems that are close to teachers' work occurs at the individual school and teacher levels. It is here that contextual and student learning data are collected, analyzed, and studied, and results reported, and so it is at these levels that evaluation expertise, capacity, and a culture that emphasizes and values learning are built. Under this set of conditions, teachers and school administrations are cast primarily in the roles of data providers and users, and, as a consequence, the public's trust can be built in the professionalism and capacity of teachers and school administrators to monitor and evaluate adequately the performance of schools and teachers, including the learning of students.

What Are Current Approaches to Assessing Learning and Their Evaluative Potential?

Having discussed the historical and professional foundations for using student performance in teacher evaluation, and having described two different approaches to doing so, we turn now to an important set of issues that must be addressed if evidence of student learning is to be included in a teacher evaluation system. These issues center on the appropriateness and value of various approaches to assessment when measures of student learning are used for different purposes. These issues can be summarized by the following questions: Which types of assessment of student achievement should be included as evidence of teachers' effectiveness? And, importantly, Why?

To answer these difficult questions, we now turn to an overview of technical considerations surrounding the collection and use of student performance data as part of a broader teacher evaluation system. These include (a) the formative and summative purposes of teacher evaluation, (b) commonly available measures of student performance, and (c) criteria for appraising measures of student performance. To summarize these issues we provide a framework, based on purposes for evaluation and criteria for appraising student measures, that can be helpful for decision making regarding which measures of student learning to include in a broadly conceived teacher evaluation system.

Formative and Summative Evaluation

We start this discussion by making a distinction between the *formative* and *summative* purposes of evaluation, because their technical requirements differ (Millman, 1981) and they should never be confused—though they can and, in our view, must support each other. As defined in Wheeler, Haertel, and Scriven's *Teacher Evaluation Glossary* (1993), *formative* teacher evaluation is the act of identifying strengths and weaknesses for the purpose of professional development, whereas *summative* evaluation is the act of making personnel decisions about a teacher for purposes of promotion, dismissal, tenure, and the like. Alternative language frequently used is evaluation for *improvement*, in the case of formative evaluation, and *accountability*, in the case of summative evaluation.

One of the principles for any teacher evaluation system, and especially those that include evidence on student performance, is that it should serve both improvement and accountability purposes at both the individual and school levels. Stronge (1994) has provided an excellent discussion of why organizational and individual goals should support each other so that the evaluation of teachers not only serves individual improvement but also advances the mission of the school. We support this view and also contend that no single approach to the assessment of student learning can adequately provide the information needed to serve all the evaluative purposes desired. Schools today employ a wide variety of student assessment instruments and procedures, and in order to bring efficiency, usefulness, and effectiveness to the collection and use of student performance data for evaluative purposes, teacher evaluation systems should be integrated into a broader program or school evaluation system (Cizek, 1995; Stronge, 1994; Webster, 1994). This view is consistent with the concept of planned assessment systems described by Cizek and his colleagues (Cizek, 1995; Cizek & Rachor, 1994).

Common Measures of Student Performance

Cizek (1995) describes the kinds and levels of student assessment commonly taking place in schools today. From his perspective, the problem faced is not a paucity of information but rather a fractured picture of how the various pieces and parts that make up the student assessment puzzle fit together in any purposeful manner.

Student performance measures commonly used within schools can be classified as follows: standardized achievement tests (national); curriculum-aligned standardized achievement tests (state or district); performance-based assessments (state or district); teacher-made objective tests; teacher-developed performance-based assessments; and observations, peer appraisals, and self-reports (Linn & Gronlund, 1995). Although this list is not exhaustive of the potential sources of information about student performance, it does provide a sense of the scope and range of student performance data already available to schools and teachers.

Criteria for Student Performance Measures

If evidence on student learning is to be used for purposes of teacher and school evaluation, measures used to collect this evidence must meet certain technical and ethical criteria. Millman's (1981) criteria for appraising student performance measures used in formative and summative evaluation serve as important guides in this regard. Additionally, volumes have been written on the development of "good" student assessment measures (see, for example, Airasian, 1994; Linn & Gronlund, 1995; Thorndike, Cunningham, Thorndike, & Hagen, 1991). The purpose here is not to review all that is known about good measurement, but rather to discuss generally the technical/ethical criteria that need to be considered in the appraisal of student measures, and then to sort various types of student measures against these criteria.

Free from bias. Any measure of student performance, whether used for formative or summative teacher or school evaluation, should be free from bias. That is, the score(s) produced by the measure should not be the result of personal preference or inclinations; rather, they should be the result of the employment of objective, fair, and impartial scoring rules (Wheeler et al., 1993). Other forms of bias commonly discussed within the context of measuring student performance include issues of disability, culture, and race and/or language, all of which may require some adaptation, either in the actual measure or in its administration, to ensure objectivity, fairness, and impartiality. It is probably the case that total freedom from bias is impossible to achieve, so the issue is to reduce the level of bias to a minimum. This criterion is not in conflict with other criteria, nor has it been cited as a cause of failure in past efforts to use student performance in teacher evaluation.

Useful for diagnosis. Although measures used for summative teacher or school evaluation do not need to be useful for diagnosis, measures used for formative purposes, by definition, must be. This criterion is the most difficult to achieve across evaluative purposes and foci in that there are both a time element and a focus element involved in defining what *useful for diagnosis* means. In terms of time, there is an immediate time frame for teachers to alter instruction during the course of a lesson, day, or unit. This is a different issue from dealing with an appropriate direction for professional development over a summer or year. At the same time, what it means to be useful for diagnosis may differ for a teacher and a school. If the position is taken that a school is simply the sum of its teachers (Sanders & Horn, 1994), then this is not an issue. If, on the other hand, a school is considered more than this, if it is considered to have a unique culture and personality of its own, then this issue does pose a problem, and it may be necessary to use different measures to satisfy these different purposes. Also, usefulness for diagnosis may be in conflict with the criterion of being broadly conceived, at least if resources and time are of critical importance. Finally, this criterion may also be in conflict with the criterion of comparability. Past efforts to use student performance in teacher evaluation have been criticized on this basis.

Valid. Any measure of student performance, whether used for formative or summative evaluation, should be sensitive to (be able to detect) the impacts of what teachers and schools do; that is, measures of student learning should have instructional validity. If they do not, but rather are measures of general intelligence or of learning that takes place outside of school, then it would be hard to justify their use for either teacher or school evaluation of any kind. This aspect of validity is not in conflict with any other criterion, but prior attempts to use student learning when evaluating the performance of teachers have been criticized on this basis.

A second aspect of validity that must be considered is curricular validity. Any measure of student performance, whether used for formative or summative teacher or school evaluation, should be consistent with the curricula of courses, programs, and/or schools. The measures should reflect both the scope and complexity of the content taught. If they do not, then it would be hard to defend the claim that a full and representative sample of teachers' or schools' work is reflected by student performance data. Worse, it may be that the student performance measures assess content that *is not* part of the curricula of courses, programs, and/or schools. This would be akin to holding teachers and schools accountable for outcomes for which they are not responsible. This aspect of validity is not in conflict with any other criterion, but, similar to instructional validity, prior efforts to use student performance in teacher evaluation have been criticized on this basis.

Reliable. Any measure of student performance, whether used for formative or summative teacher or school evaluation, should be reliable. That is, the measure should produce consistent results (a) across time for the same set of students performing at the same levels, (b) across time for the same scorer(s) using the measure with the same students, and (c) across scorers using the measure for the same students at the same time (Wheeler et al., 1993). Perfect consistency is rarely achieved, so the question becomes, How reliable is reliable enough? The answer to this question depends on the use, or the decisions made based on the scores (Linn & Gronlund, 1995). The more important the decision, and the less reversible the decision in the future, the higher the level of reliability required. For example, a measure used to decide if additional instruction is needed requires a much lower level of reliability than a measure used to make placement or status decisions that are not easily reversible. This criterion is not in conflict with others, nor has it been cited as a cause of failure in past efforts to use student performance in teacher evaluation.

Broad. Teachers and schools are responsible for fostering a broad array of outcomes, some of which go beyond student learning. Similarly, teachers and schools are responsible for fostering student learning that goes far beyond the acquisition of subject-matter knowledge. As a result, any measure of student performance, whether used for formative or summative purposes, should be broadly conceived. This may be in conflict with the criterion of being useful for diagnosis. Prior efforts to use student performance in teacher evaluation have been criticized on this basis in terms of narrowing the curriculum.

Equitable (fair). This criterion applies directly to the situation in which all teachers or schools are evaluated on the basis of student performance on a common measure. In this case, no teachers or schools that make up a comparable group (e.g., all first-grade teachers in a building) should be placed at either an advantage or a disadvantage compared with any other based on the actual measure. This criterion applies most directly to the use of student performance data for summative evaluation. Prior efforts to use student performance in teacher evaluation have been criticized on this criterion, primarily because no one measure can meet all other criteria for all teachers working at a single grade level, let alone a school.

Comparable. One aspect of being equitable (fair) is that results for one teacher or school should be comparable to results for other teachers and other schools. No teacher or school should be disadvantaged compared with any other based on factors beyond their control. Haertel (1986), Berk (1988), and Millman and Sykes (1992) all have provided thoughtful analyses of the issues related to equity and comparability. Suffice it to say that it is this issue that has scuttled many previous efforts to include evidence of student performance in teacher evaluation. No measure can produce scores that are comparable across teachers or schools until the scores have been adjusted to take differences in contexts (students, classrooms, schools, communities) into account.

Of the seven criteria just discussed, four are necessary for assessments serving any purpose: freedom from bias, validity (instructional and curricular), reliability, and breadth of coverage. Although these are necessary criteria, they are not sufficient. To be useful for formative (improvement) evaluation, measures also must be useful for diagnostic purposes. To be useful for summative (accountability) evaluation, measures must also be equitable and comparable.

A Framework for Choosing Measures of Student Performance in Teacher Evaluation

Having outlined uses, measures, and criteria, we now need to determine how commonly available measures of student learning fare against our criteria. Table 7.2 provides some initial appraisals in this regard. It is important to note, however, that our appraisals have been made on the basis of reality rather than ideals. Thus, for example, although teachers *can* construct tests to be reliable and valid, current levels of "assessment literacy" for most teachers (Stiggins, 1991, 1995) make this the exception rather than the rule, indicating that student assessment is a major area of need for teachers' professional development.

Our appraisals of commonly used measures of student performance against multiple criteria of quality lead us to conclude that a silver bullet measure does not exist. Two glaring weaknesses show up in Table 7.2. The first is that no single assessment approach meets the criterion of being broadly conceived. Although it is true that an entire battery of national or state standardized tests comes close to meeting this criterion, it still falls short for any one teacher or school. Second, no measure can produce scores that are strictly comparable across teachers or schools. This would require that all of the factors that influence student performance that lie beyond the control of a

TABLE 7.2 An Appraisal of Commonly Available Measures of Student Performance for Use in Evaluative Systems

	Evaluation Uses			
Assessment Measures	*Individual Improvement*	*Program/School Improvement*	*Individual Accountability*	*Program/School Accountability*
Nationally normed tests of achievement	freedom from bias; reliability	freedom from bias; reliability; fairness	freedom from bias; reliability	freedom from bias; reliability; fairness
State-normed curriculum-aligned tests of achievement	freedom from bias; reliability; instructional validity; curricular validity; useful for diagnosis	freedom from bias; reliability; useful for diagnosis; instructional validity; curricular validity	freedom from bias; reliability; instructional validity	freedom from bias; reliability; fairness; instructional validity; curricular validity
State-developed performance assessments	instructional validity; curricular validity; useful for diagnosis	instructional validity; curricular validity; useful for diagnosis	instructional validity; curricular validity	instructional validity; curricular validity; fairness
Teacher-developed objective tests	useful for diagnosis; instructional validity; curricular validity	instructional validity; curricular validity	instructional validity; curricular validity	instructional validity; curricular validity
Teacher-developed performance assessments	useful for diagnosis; instructional validity; curricular validity	instructional validity; curricular validity;	instructional validity; curricular validity	instructional validity; curricular validity

teacher or school be randomly distributed across all teachers and schools. This condition can be satisfied only if the effects of these factors are statistically controlled.

In reflecting upon the multiple purposes of evaluation to which student performance can be put, and reviewing Table 7.2, we can conclude that Shulman's (1988) notion of teacher assessment as a "union of insufficiencies" can also be applied to measures of student performance. No single kind of assessment rates high on more than five criteria for any one purpose. This does not mean that student performance cannot be used for evaluative purposes, only that no one measure will suffice. Just as multiple lines of evidence should be part of any good evaluation system (McConney, 1994b), so too should multiple measures of student performance. Table 7.3 presents a simple framework that may be useful for thinking about how different sources (measures) of student performance currently available in a school may serve one or more purposes in a broadly conceived teacher evaluation system. The table lists the criteria that are most important for given evaluative purposes, as well as several kinds of assessment that best meet the criteria. For example, in the first column and first row, the table describes individual teacher improvement, with important criteria including usefulness for diagnosis and instructional and curricular validity. These criteria are likely best met by a combination of student measures, including teacher-developed objective and performance-based assessments, state-developed curriculum-aligned assessments, and state-developed performance assessments.

In the context of Table 7.3, when deciding on which measures to include in a teacher appraisal system, the first step for a school to take is to reach consensus within its faculty about how well each available measure of student learning stacks up against each criterion in Table 7.2. It is likely that each school/district will bring a somewhat unique set of circumstances against which to make these judgments. In addition, other student performance measures available in a school or district should be examined for possible inclusion. At the secondary level, these might include measures such as advanced placement exams; band, speech, drama, and Future Farmers of America performance requirements; and debate competitions that provide some measure of student performance in areas not commonly mea-sured on state or national exams. At the elementary level, the list is probably much shorter.

TABLE 7.3 The Multiple Purposes of Evaluation: Matching Measures to Uses

	Uses	
Focus	*Improvement (Formative Evaluation)*	*Accountability (Summative Evaluation)*
Individual teacher evaluation	A combination of measures that are, most importantly, useful for diagnosis and have both instructional and curricular validity for teachers. For example: (a) teacher-developed objective tests (b) teacher-developed performance assessments (c) state curriculum-aligned achievement tests (d) state-developed performance assessments	A combination of measures that are, most importantly, valid (curriculum and instruction), reliable, unbiased, broadly conceived, and fair for teachers. For example: (a) state curriculum-aligned achievement tests (b) state-developed performance assessments (c) nationally normed achievement tests (d) teacher-developed objective tests (e) teacher-developed performance assessments (f) observations, peers reviews, self-reports
Program school or system evaluation	A combination of measures that are, most importantly, useful for diagnosis and have both instructional and curricular validity for schools. For example: (a) state/district curriculum-aligned achievement tests (b) state/district-developed performance assessments	A combination of measures that are, most importantly, valid (curriculum and instruction), reliable, unbiased, broadly conceived, and fair for schools. For example: (a) state curruiculum-aligned achievement tests (b) nationally normed achievement tests (c) state-developed performance assessments

What Is the Proper Balance in Using Student Learning in the Evaluation of Teachers?

By now, as we begin the capstone section of this chapter on using student achievement data for teacher evaluation, the reader will have noted a recurring theme in our discussions. That theme is *balance*. Over time, our society, and especially that part of it directly involved in our educational systems, has recognized a need for balance in the assessment and evaluation of children's academic growth and accomplishments. We would be less than satisfied, for instance, if a yearly report of a child's academic growth and accomplishments was based solely on one 45-minute period of observation per year. We would be similarly dissatisfied if the report included the child's accomplishments in athletic and social activities but made no mention of achievement in academic learning. It is time for the same good principle of balance that we apply to the evaluation of our students to be applied to our teachers and educational institutions. Balance, especially the sustainable variety, is essential in the design of sound evaluation systems for teachers and facilitates (or even demands) building a school culture where the inclusion of measures of student growth in learning is viewed as the *cornerstone source* of evaluative data.[7]

Principles of Teacher Evaluation System Design

How do we design and build a teacher evaluation system that includes student learning and is at the same time balanced and sustainable? McConney (1994a) has noted six general guiding principles of evaluation system design in synthesizing the views of researchers at the Center for Research on Educational Accountability and Teacher Evaluation (CREATE). These principles require the following:

1. Linking the evaluation system to the goals and needs of the school in which a teacher works as well as to the goals and needs of each teacher within the school
2. Linking the evaluation system to the duties and responsibilities designated for a teacher by the profession (we include in these the fostering of desired learning outcomes), by a contract, and by a job definition

3 and 4. Agreeing prior to evaluation on the multiple lines of evidence to be considered in the evaluative process (we include in these evidence of student progress in learning by students taught) and the performance standards to be met around each line of evidence for one's work to be judged *exemplary*, *satisfactory*, or *unsatisfactory*

5. Agreeing prior to evaluation that the evaluative process is to serve formative (improvement) as well as summative (accountability) purposes, that everyone involved will know clearly which purpose is being served prior to a particular evaluative cycle, and that everyone involved will assess the evaluative data collected from the perspective of *institutional* as well as *individual* improvement
6. Holding the evaluation system designed up to the criteria outlined in the Personnel Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1988) and modifying the system as needed to meet the standards proposed for propriety, utility, accuracy, and feasibility

Although these principles are both sensible and sound, they overlook sources of tension that are difficult to resolve, as they assume the accomplishment of three critical tasks that require attention when including evidence of learning gains by students taught as a part of teacher evaluation. The most obvious are the following assumptions:

- That a workable balance has been achieved between the goals and needs of a school and the goals and needs of teachers working within a school
- That a workable balance has been achieved between the formative (improvement) and summative (accountability) purposes to be served by the evaluation system
- That a workable balance has been achieved between the costs (time and other resources) required for teachers and principals to implement such a system fully and the benefits to be gained

Recommendations

Finding balance within and among these potential sources of tension is a process that requires ongoing, open communication and collaboration among all stakeholders in the teacher (and school) evaluation system. To bring resolution to the potential sources of tension noted above, and to help accomplish the sustainable approach we advocate, we strongly recommend striving for the following kinds of balance.

Balance in the sources and types of data applied in the evaluation of teachers. There are very few, if any, measurement methodologies that by themselves can or should carry the entire burden of personnel decisions.

> The necessity for using multiple sources of data in the evaluation of school professionals . . . is a basic and central principle of educational measurement in that any one data source or instance of measurement is simply *one sample of behavior*, and the greater the variety and number of samples taken the better (more reliable) the representation of performance over time. (McConney, 1995, p. 107)

Student achievement data should be an important source, but only one of a number of sources, because even though we view student learning as the primary responsibility of teachers and schools, teachers do have other professional duties and responsibilities for which they are accountable (e.g., Scriven, 1988, 1994).

Balance in the types of student achievement data used for the evaluation of teachers. As we have discussed in a previous section, no one type of student assessment adequately meets all the criteria necessary to support its inclusion in a system of teacher evaluation that serves formative and summative purposes. Further, no one type of student assessment adequately describes the full range of learning we increasingly recognize as essential for our students to be successful citizens in the 21st century. For these reasons we advise and advocate the inclusion of multiple measures of student learning in systems that include such information for teacher or school evaluation purposes.

Balance in the use of assessment/evaluation data for formative and summative purposes. Although students and parents who support and use our schools must be assured of appropriate services, teachers also

BEST COPY AVAILABLE

must have the opportunity to improve, as this will foster support and ownership of the evaluation system and the school's improvement objectives and facilitate the improvement and capacity building of a school's personnel. As Stufflebeam (1983) has pointed out, the primary purpose of evaluation is "not to prove but to improve" (p. 117). We share this view wholeheartedly, and we believe that balance must, if anything, weigh heavily in favor of providing authentic opportunities for diagnosis (needs assessment), improvement, professional development, and capacity building for teachers.

Balance between centralized and teacher-centered, school-based monitoring and evaluation. There are costs and benefits associated with the two approaches to teacher evaluation (centralized and school based) we have described, and these need to be weighed in the design of any teacher evaluation system. For instance, whereas centralized systems using large-scale, objective student assessments typically are more psychometrically defensible, they seem tilted toward building evaluation capacity and expertise at the district or state level. On the other hand, although less psychometrically rigorous, school-based teacher evaluation systems anchored to teacher- or school-made student assessments are closer to teachers' work, and they encourage evaluation capacity building at the teacher and school levels. A mix of these two orientations would seem most sensible and most likely would enhance the professionalization of teaching by building the public's trust in teachers and schools, in addition to trust in the central office staff that supports them.

Summary

The available evidence (philosophical, ethical, rational, psychometric, technical-methodological, and political-contractual) leads us to conclude that the use of student performance data for teacher evaluation is entirely possible and defensible. Such information, however, should never form the sole basis for making judgments about teachers' effectiveness.

Overall, at the system level, we suggest three monitoring-evaluation orientations that we believe are becoming more accepted in concept, prevalent in practice, and acknowledged by both evaluation specialists and school professionals as sensible for the evaluation of educational professionals, including teachers:

- Multifaceted data collection at both student and teacher levels
- Evaluation for improvement and informed decision making
- Open and ongoing communication and collaboration among all stakeholders in the design and operation of the evaluation system

These three orientations, in conjunction with the principles and recommendations noted earlier, should facilitate the design of a teacher evaluation system that uses student performance as an important source of data and ensure that the system is both balanced and sustainable.

Notes

1. Many merit pay and career ladder evaluation systems for teachers require evidence of student learning, but this is not the case in most regular teacher evaluation systems operating in the schools (Berk, 1988; Murnane & Cohen, 1986; Smylie, 1994).
2. The discussion in this section is adapted from Millman and Schalock (in press).
3. This section is not intended as a comprehensive description of every educational system or subsystem that makes use of student performance data for teacher evaluation. We have chosen to use as examples systems in Dallas (Texas), Tennessee, and Oregon because, in different ways and to varying degrees, these systems have overcome many of the methodological challenges that previously have prevented the defensible use of student achievement data in teacher evaluation, and we believe that these examples provide useful variability in approach when viewed against the taxonomy in Figure 7.1. We also point out that the continua in Figure 7.1, although somewhat similar in concept to the distinction between formative (developmental) and summative (accountability) perspectives in education personnel evaluation, more accurately represent our judgments about who bears the professional costs (e.g., time, resources, effort) and reaps the professional and capacity-building benefits (e.g., assessment/evaluation skill, personnel and institutional alignment, large-scale data management and reporting) that go along with varying approaches to teacher evaluation. Further, we reemphasize that our focus here is on the use of student achievement data as a criterion for *teacher performance evaluation* (at pre- and in-service levels) and so will not attend to the use of student achievement for teacher or school incentive programs. We make this distinction because, as noted, our descriptions are illustrative, not comprehensive, and according to Berk (1988), there are many states (about one-third) in which the use of student performance results "for statewide teacher incentive/school

incentive/career ladder programs" is under exploration or implementation (p. 346).

4. It is important to note that of the three approaches described here, only that of the Dallas Public Schools is currently fully functional for the performance evaluation of *practicing teachers*. Some aspects of the Dallas system are still under field test. For instance, a number of groups of teachers (about 50% of all teachers) are not assigned effectiveness indices due to the unavailability of objective student achievement assessments for certain subjects and grades (e.g., middle school science, visual arts, music, physical education) or because of unusually low numbers of students (fewer than 11) on whom to base the effectiveness indices. Teachers without effectiveness indices are assessed using an alternative instrument, the Assessment of Teacher Accomplishments and Performance (ATAP). The conceptual basis for the items and scoring of this integrated assessment is consistent with that undergirding Classroom (Teacher) Effectiveness Indices; that is, teachers have great influence on the academic growth of their students, and, indeed, academic growth is the primary responsibility of the teacher. The evidence required of teachers by this performance assessment, therefore, includes evidence of growth in student learning (on a student-by-student basis). Further, the scoring procedure for the assessment reflects the preponderant emphasis that Dallas Public Schools and the Dallas community have placed on the academic growth of students, as is evident in Texas Senate Bill 1.

   In Tennessee, the state's Educational Improvement Act, which took effect July 1, 1992, called for the system to report individual teacher effects on students' academic growth to "teachers, appropriate administrators, and school board members no later than July 1, 1995" (Sanders & Horn, 1992, p. 2). However, this target was postponed so that the soundness of the TVAAS could be externally evaluated by a panel of experts. The TVAAS was endorsed as technically sound in this evaluation; the panel's report is contained in what has come to be known as the Bock Report. Currently, data resulting from the TVAAS are used to gauge the effectiveness of schools, although the system is designed for and does have the capacity to disaggregate school effects to the level of individual teachers. Teacher effects were scheduled to be reported in fall 1996.

5. In Oregon, the requirements associated with teacher work sample methodology are currently fully in place as part of the state's initial and continuing teacher licensure system (both as part of Oregon's Teacher Standards and Practices Commission teacher licensing policy and as an integral component of teacher education programs around the state), but this system does not yet extend to the regular performance appraisal of teachers.
6. Teacher work sample methodology has, however, attracted criticism in that rigorous evidence for the psychometric defensibility (e.g., adequate reliability) of the approach is lacking.

7. Before turning to a discussion of how balance manifests itself in the design of a teacher evaluation system that includes student achievement data, we need to acknowledge two other issues: the contractual (legal) and political issues surrounding the use of student learning data in teacher evaluation. Although we recognize that these issues are undoubtedly real, complex, and intertwined, we also recognize that they are very dependent on the community, district, state, and national contexts in which a school operates. Therefore, we will not address these issues directly except to advise that open communication and collaboration among all stakeholders should be the hallmark of the design and operation of any sound system of evaluation. This will foster ownership of the system and facilitate the inclusion of student learning data as part of teacher and school evaluation.

References

Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.

Bembry, K. (1996). *The new Dallas teacher evaluation system: After the first year.* Paper presented at the Fifth Annual National Evaluation Institute, Bethesda, MD.

Berk, R. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education, 1*, 345-363.

Cizek, G. (1995). The big picture in assessment and who ought to have it. *Phi Delta Kappan, 77*, 246-249.

Cizek, G., & Rachor, R. (1994, March). The real testing bias: The role of values in educational assessment. *NASSP Bulletin*, pp. 83-93.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. *Educational Evaluation and Policy Analysis, 8*, 45-60.

Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards: How to assess systems of evaluating educators.* Newbury Park, CA: Sage.

Linn, R., & Gronlund, N. (1995). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

McConney, A. A. (1994a). Preface. In A. A. McConney (Ed.), *Toward a unified model: The foundations of educational personnel evaluation* (pp. i-v). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

McConney, A. A. (Ed.). (1994b). *Toward a unified model: The foundations of educational personnel evaluation*. Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

McConney, A. A. (1995). Introduction: Common ground for a unified approach. *Studies in Educational Evaluation, 21*, 105-110.

McConney, A. A., & Schalock, M. D. (1996). Teacher work sample methodology: Promise for program *and* personnel appraisal? *AASPA Report*, 3(8), 16.

Millman, J. (1981). Student achievement as a measure of teacher competence. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 147-166). Beverly Hills, CA: Sage.

Millman, J., & Schalock, H. D. (in press). Beginnings and introduction. In J. Millman (Ed.), *Assuring accountability . . . ? Using gains in student learning to evaluate teachers and schools*. Newbury Park, CA: Corwin.

Millman, J., & Sykes, G. (1992). *The assessment of teaching based on evidence of student learning: An analysis* (Research Monograph No. 2). Washington, DC: National Board for Professional Teaching Standards.

Murnane, R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review, 56*, 1-17.

Popham, W. J. (1971). Performance tests of teaching proficiency: Rationale, development, and validation. *American Educational Research Journal, 8*, 105-117.

Sanders, W. L., & Horn, S. (1992). *An overview of the Tennessee Value-Added Assessment System (TVAAS), with answers to frequently asked questions*. Knoxville: University of Tennessee, Statistical and Computing Services.

Sanders, W. L., & Horn, S. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-312.

Sanders, W. L., & Horn, S. (1995). Educational assessment reassessed: The usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes. *Education Policy Analysis Archives* [On-line], 3(6). Available (gophered) at INFO.ASU.EDU in the subdirectory CAMPUS-WIDE INFORMATION.

Schalock, H. D., Schalock, M. D., & Girod, G. (in press). Linking learning gains to teacher work: Oregon's work sample methodology. In J. Millman (Ed.), *Assuring accountability . . . ? Using gains in student learning to evaluate teachers and schools*. Newbury Park, CA: Corwin.

Schalock, H. D., Schalock, M. D., Myton, D., & Girod, G. (1993). Focusing on learning gains by pupils taught: A central feature of Oregon's outcome-based approach to the initial preparation and licensure of teachers. *Journal of Personnel Evaluation in Education, 7*, 135-158.

Scriven, M. (1988). Duties-based teacher evaluation. *Journal of Personnel Evaluation in Education, 1*, 319-334.

Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education, 8*, 151-184.

Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*(3), 36-41. (ERIC Document Reproduction Service No. EJ 385 344)

Smylie, M. A. (1994). Redesigning teachers' work: Connections to the classroom. In L. Darling-Hammond (Ed.), *Review of research in education* (pp. 129-178). Washington, DC: American Educational Research Association.

Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan, 73*, 534-539.

Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*, 238-245.

Stronge, J. (1994). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. In A. McConney (Ed.), *Toward a unified model: The foundations of educational personnel evaluation* (pp. 17-34). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services in evaluation* (pp. 117-141). Boston: Kluwer-Nijhoff.

Teacher Evaluation Task Force, Dallas Public Schools. (1995). *Manual for the proposed teacher evaluation system*. Dallas: Author.

Teacher Evaluation Task Force, Dallas Public Schools. (1996, April 8). *Revised manual for the proposed teacher evaluation system* (Draft). Dallas: Author.

Teaching Research Division, Western Oregon State College. (1995, October 5). *Teacher effectiveness in promoting learning: An overview of research in progress at Western Oregon State College* (Report prepared on request for the Oregon Governor's Office). Monmouth. OR: Author.

Thorndike, R. M., Cunningham, G., Thorndike, R. L., & Hagen, E. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York: Macmillan.

Webster, W. J. (1994). The connection between personnel evaluation and school evaluation. In A. A. McConney (Ed.), *Toward a unified model: The foundations of educational personnel evaluation* (pp. 93-117). Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Webster, W. J. (1995). The connection between personnel evaluation and school evaluation. *Studies in Educational Evaluation, 21*, 227-254.

Webster, W. J., & Mendro, R. L. (1995, January). *Evaluation for improved school level decision-making and productivity*. Paper presented at the Hawaii Institute on Assessment and Accountability, Honolulu.

Webster, W. J., Mendro, R. L., Bembry, K. L., Orsak, T. H., & Weerasinghe, D. (1995, October). *Alternative methodologies in the development of teacher and school effectiveness indices.* Paper presented at the annual meeting of the Rocky Mountain Educational Research Association, Albuquerque, NM.

Wheeler, P. H. (1994). Before you use portfolios in teacher evaluation . . . consider these issues. *AASPA Report, 1*(4), 8-9.

Wheeler, P. H., Haertel, G., & Scriven, M. (1993). *Teacher evaluation glossary*. Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

Wolf, K. (1994). Teaching portfolios: Capturing the complexity of teaching. In L. Ingvarson & R. Chadbourne (Eds.), *Valuing teachers' work: New directions in teacher appraisal* (pp. 112-136). Melbourne: Australian Council for Educational Research.

8

Portfolios in Teacher Evaluation

KENNETH WOLF
GARY LICHTENSTEIN
CYNTHIA STEVENSON

Teaching portfolios are increasingly popular tools for both evaluation and professional development. The use of portfolios is currently in place in a variety of settings, including in university education programs, to foster the professional growth of preservice teachers (Borko, Michaelec, & Timmons, 1996); in school districts, to reward outstanding practice (Wolf, Lichtenstein, Bartlett, & Hartman, 1996); at the state level, for license renewal (Lichtenstein, Wolf, Pease, Ruckle, & Campbell, 1996); and at the national level, to certify accomplished practitioners (National Board for Professional Teaching Standards, 1989).

Why have educators turned to teaching portfolios? Proponents contend that portfolios present authentic views of learning and teaching over time, offering a more complete and valid picture of

---

AUTHORS' NOTE: We would like to thank Janet Junkin and Nancy Hall for contributing examples from their teaching portfolios for this chapter.

what teachers know and can do (Shulman, 1988; Wolf, 1991, 1996). Moreover, they believe that portfolios promote professional development by providing teachers with a structure and process for documenting and reflecting on their practice.

Although portfolios have many attractive features, their use can have significant liabilities as well. Portfolios can be time-consuming to construct, cumbersome to store, and difficult to score. Nevertheless, the potential of portfolios for addressing assessment needs as well as advancing professional learning suggests that administrators should consider a role for teacher portfolios in their school-based evaluation and staff development programs.

In this chapter we provide principals and other school administrators with an overview of teaching portfolios so that they can make informed choices about their use. We begin by defining a teaching portfolio and discussing issues related to their design. Next, we present examples from actual teaching portfolios. We conclude by offering recommendations for putting portfolios into practice.

What Is a Teaching Portfolio?

In its most basic form, a teaching portfolio is a collection of information about a teacher's practice. Portfolios can come in many different shapes and sizes, but in practice they often take the form of scrapbooks filled with photographs of classroom life, along with affectionate notes from students and parents. Although this kind of portfolio may be eye-catching and heartwarming, it does not effectively advance either evaluation or professional development goals.

The problem with portfolios such as these is that they are not connected to professional content standards that describe what teachers should know and be able to do, nor are they explicitly linked to a teacher's individual teaching philosophy or a school's improvement plan. Additionally, these types of portfolios often lack examples of student or teacher work that illustrate the ways in which the teacher has acted on the professional content standards and their philosophies and goals. Moreover, there is no explicit reflection on the teaching and learning portrayed in such portfolios, nor explanation of the context in which these events occurred. Furthermore, these portfolios are typically constructed without input from colleagues.

What might be a more productive vision for portfolios? Drawing on the work of Lee Shulman (1992), we offer the following:

> *A teaching portfolio is the structured documentary history of a carefully selected set of coached or mentored accomplishments, substantiated by samples of student work, and fully realized only through reflective writing, deliberation, and serious conversation.*

The key features of such a teaching portfolio are as follows:

- The portfolio should be *structured* around sound professional content standards and individual and school goals.
- The portfolio should contain *carefully selected examples of both student and teacher work* that illustrate key features of the teacher's practice.
- The contents of the portfolio should be framed by *captions* and written *commentaries* that explain and reflect on the contents.
- The portfolio should be *a mentored or coached experience*, in which the portfolio is used as a basis for ongoing professional conversations with colleagues and supervisors.

What Purposes Might a Teaching Portfolio Serve?

There are three main purposes for preparing a teaching portfolio:

- To address evaluation requirements
- To advance professional growth
- To aid in employment searches

Although a carefully conceptualized portfolio can address all of these to some degree, each purpose suggests somewhat different design considerations. Portfolios that are used primarily for evaluation, for example, require greater structure than do those that primarily serve professional growth. For the evaluation portfolio, *fairness* is a chief concern. Consistency—in portfolio requirements and in the evaluation process—best advances this goal. With the professional development portfolio, however, *ownership* of the learning process is a major concern. Individual customization of the portfolio best serves this goal. For an evaluation portfolio, for example, the structure and contents should be specified in advance so that the requirements for completing the portfolio are clear and the evaluation process is consistent. On the other hand, for a professional development portfolio, latitude in choices concerning the focus and format of the portfolio is likely to increase learning, because each teacher will adapt the portfolio to his or her specific needs and goals.

A portfolio used in employment searches is shaped by still different forces. In all likelihood, those doing the hiring will have limited time to review the portfolio, and they will be strongly affected by their general impressions. In this situation, as Lichtenstein (1997) has found in his interviews with administrators, the portfolio author needs to pay greater attention to presentation issues, such as the attractiveness of the portfolio and the accessibility of the information.

Although we distinguish among these three types of portfolios in terms of their primary purposes to highlight the trade-offs inherent in each, we want to emphasize that a single portfolio can advance all three goals if the person responsible for conceptualizing the portfolio is clear about his or her purposes and thoughtful in design.

What Might Be Included in a Teaching Portfolio?

The contents of teaching portfolios can be as varied as the people who construct them. A portfolio might include samples of student and teacher work, such as photographs of class projects, lesson plans, student assessments, and evidence of professional activities. In addition, a portfolio might include a variety of other information, such as letters of recommendation from parents, evaluations from supervisors, and even teaching credentials and college transcripts. Any number of these items might be desirable, but which are essential? We recommend that the following items be included in every portfolio, regardless of its primary purpose:

- A statement of philosophy or teaching goals
- Samples of teacher work, such as lesson plans and student assessments
- Samples of student work, such as reading logs and student projects
- Captions that briefly explain the work samples
- Commentaries that reflect on the teaching and learning documented in the portfolio

In addition to samples of student and teacher work, captions and commentaries on those samples are essential. Captions provide con-

Title of evidence: ______________________________

Date created: ______________________________

Educator's name: ______________________________

Description of context in which evidence was collected: ________

____

Interpretation: ______________________________

____

Additional comments: ______________________________

____

____

Figure 8.1. Portfolio Caption Form Used by the Colorado Department of Education

textual information about each item in the portfolio (see Figure 8.1 for an example of a caption format from the Colorado Department of Education). Commentaries are written accounts that elaborate on and interpret the portfolio contents.

In essence, then, a portfolio should be framed by a sound philosophy that is consistent with professional expectations and school goals, illustrated through samples of student and teacher work, and explained through captions and commentaries. Depending upon the purposes of the portfolio, however, additional information might be warranted. For example, an employment portfolio should probably contain a résumé, whereas an evaluation portfolio might include copies of previous evaluations. Such additional contents will vary across settings based on individual and school needs; the ingredients described above, however, are essential for all portfolios.

How Might the Portfolio Process Unfold?

The following steps might serve as a guide for teachers in building their portfolios:

- Prepare a philosophy statement.
- Set goals for the portfolio in consultation with a supervisor.
- Collect a variety of student and teacher work samples.
- Discuss the work samples with colleagues at regular intervals.

- Organize and caption the portfolio contents.
- Write reflective commentaries about the teaching and learning documented in the portfolio.
- Submit the completed portfolio to a supervisor for review.
- Receive feedback from the reviewing supervisor.
- Set new goals in light of the supervisor's feedback.

In this scenario, the teacher and his or her supervisor meet to set the goals and expectations for the teacher's portfolio. Throughout the school year, the teacher gathers a variety of work samples for possible inclusion in the portfolio. At regular intervals, the teacher uses these artifacts as departure points for discussions with colleagues about his or her teaching. The teacher prepares to submit the completed portfolio by organizing and captioning the portfolio contents and explaining through written commentary the significance of various items. Finally, the portfolio is formally evaluated by the teacher's supervisor, who rates the portfolio performance and provides written, and possibly oral, feedback to the teacher. The teacher then sets new goals in light of what he or she has learned. Although action is not a part of the portfolio process per se, it is assumed that the teacher will be constantly acting on what he or she is learning to improve classroom practice and students' learning.

How Should a Teaching Portfolio Be Evaluated?

Portfolios are exciting as assessment tools because they allow teachers to represent the complexities and individuality of their teaching in great detail. They are problematic, however, for the same reasons. Each portfolio is thick and unique, making evaluation a daunting task.

Ensuring that the evaluation process is manageable and fair requires that several elements be put into place in advance, including identification of sound content and performance standards for teachers, specification of the requirements for construction of a portfolio, and design of an efficient evaluation system. The establishment of these elements will increase the likelihood that the evaluation system will successfully meet the essential requirements of validity, reliability, and utility (Stronge & Helm, 1991).

Identification of Content and Performance Standards

The evaluation of a teacher's portfolio should be based on clear content standards (what teachers should know and be able to do) and performance standards (how well they should know and be able to do it). These standards should be spelled out in advance, so that teachers have clear targets for their performance. These standards will serve to guide teachers in the construction of their portfolios as well as reviewers in their evaluations.

We recommend the identification of a small set of content standards (about three to seven), with each of the standards composed of several statements that clarify its meaning. For example, the Douglas County School District in Colorado has three content standards (e.g., assessment and instruction) for evaluating a teacher's performance, each of which is explained with five statements that operationally define the standard (see Table 8.1).

Along with content standards, performance standards need to be established as well. Performance standards address the question, How good is good enough? What level of performance is required for an "outstanding" designation, for example? Are the expectations different for beginning and experienced teachers? Determining these levels may take several exploratory efforts. During the development of these performance levels, we suggest that the emphasis of the portfolio process should be on professional growth rather than on high-stakes evaluation.

How many performance levels are desirable? Ratings can vary from satisfactory/unsatisfactory to a scale with multiple designations, such as accomplished/proficient/needs improvement. Ratings can be made at the overall performance level or for each of the content standards. We recommend a rating system that is as simple as possible to avoid the problems associated with making many fine (and probably unsupportable) distinctions about portfolio performance. Feedback to teachers, however, should be detailed and linked to the information in the portfolios, so that teachers have a clear understanding of the specific strengths and weaknesses in their performance as well as the reasons for their ratings.

TABLE 8.1 Professional Content Standards for Teachers: Douglas County School District, Colorado

---

Assessment and instruction

Outstanding educators

- Act on the belief that all students can learn
- Recognize, value, and adjust for individual differences, while maintaining an in-depth understanding of how students develop and learn
- Implement a variety of instructional strategies
- Motivate and engage students
- Draw on multiple strategies for assessing student learning and development, and clearly communicate assessment results to students and parents

Content and pedagogy

Outstanding educators

- Know the subjects they teach
- Know how to teach those subjects to students
- Know how subjects are related to one another across the curriculum
- Can relate the subjects they teach to real-world applications
- Are reflective practitioners who facilitate professional growth through self-assessment, and whose knowledge is based on both practical experience and professional literature

Collaboration and partnership

Outstanding educators

- Collaborate with other school professionals
- Work effectively with parents and community
- Draw on school and community resources to benefit students
- Contribute to the school, community, and profession in a variety of ways
- Respect diverse individuals and groups

Designing the Portfolio

To help ensure that the portfolio construction and review process is manageable, a portfolio should be focused on a few key areas of teaching rather than the entire curriculum, and should be slender in size rather than as thick as a metropolitan phone book. For example, an elementary school teacher might address aspects of only two content areas, such as mathematics and art, and link the two through a similar topic, such as patterns. Along the same lines, a secondary school teacher might choose two broad topics within his or her subject matter. For example, a middle school English teacher might document aspects of his or her writing instruction and literature discussions over a semester, and a high school American history teacher might focus on units on the American Revolution and the civil rights movement.

An alternative approach to focusing on a specific topic or subject matter might be for teachers at any level to conduct small-scale case studies on three diverse learners in their classrooms in which they investigate the students' learning across a variety of content areas and contexts. The point here is that the focus for the portfolio can productively vary, but trying to address too broad a scope within a single portfolio is more often a problem than is too narrow a focus.

Less is more, not only in terms of breadth of coverage but also in amount of information. A carefully selected collection of evidence can be used more productively by both teacher and administrator than can a file cabinet's worth of material. In any case, a portfolio already contains far more information than is available in most evaluation contexts. But how much is enough? Five to ten teacher work samples, such as lesson plans and classroom tests, and a similar number of student work samples, such as homework assignments and self-assessments, might be sufficient, assuming they are carefully selected to illustrate their connection to the content standards and portfolio goals. Add a caption for each piece of information, and one or two commentaries of two to three pages in length, and the portfolio evaluator has a wide range of information on which to base decisions and feedback.

Specification of Requirements

Another way to make the portfolio construction and evaluation process more manageable and fair is to specify the requirements for the portfolio in advance. This information might be packaged in a "portfolio construction handbook," which should include the following elements:

- Purposes for the portfolio
- Procedures for constructing the portfolio
- Timeline for completion and evaluation of the portfolio
- List of required and/or suggested portfolio contents
- Description of the evaluation process
- Evaluation criteria (content and performance standards)
- Description of the feedback and appeals process

As this list indicates, teachers should be provided with specific information about how to construct portfolios, including advice on the types and numbers of work samples that should be included and the length and structure of written commentaries. Timelines for submitting materials for review by mentors or supervisors should be explicit. In addition, the evaluation process should be described, so that teachers know who will participate in their reviews, what additional sources of information might be used, and what options teachers have for revising or remedying less-than-proficient performance appraisals. Clearly defined portfolio procedures allow teachers to spend more time reflecting on their instruction and less on trying to figure out how to "play the portfolio game."

Evaluating the Contents

Given the challenge of comprehensively reviewing a thick set of documents and materials, we recommend that administrators examining teacher portfolios follow a systematic review process that includes the following steps:

- Read the entire portfolio to get a sense of the overall performance.
- Review the portfolio in light of the content standards and teacher goals.
- Take notes about significant pieces of information in the portfolio.
- Assign a rating for the portfolio (if appropriate).
- Provide feedback to the teacher.

It is important that the reviewer examine the portfolio for each of the content standards. This can best be achieved by reading the entire portfolio first, and then reviewing the portfolio with each standard as a separate lens. During these cycles through the portfolio, the reviewer should note significant sources of evidence that will be used in the evaluation and as feedback to the teacher.

Additionally, the reviewer should keep the teacher's goals for the portfolio in mind. A portfolio, although a thick collection of information, is still only a thin slice of a teacher's entire performance. Thus, it is unrealistic to expect a teacher to represent all that he or she knows and does within a single portfolio. The teacher's goals for the portfolio, which have been set in consultation with a supervisor, should provide the focus for the evaluation.

Whether the portfolio actually receives a rating (e.g., exemplary, proficient, unsatisfactory) will depend on the purposes of the program. In most cases, it may be unnecessary, and even undesirable, for the supervisor to rate the portfolio. Assigning scores to portfolios may focus conversations among teachers, and between teachers and administrators, on scores rather than on issues of teaching effectiveness. To offset this possibility, we recommend basing teacher evaluation on a variety of evidence—such as direct observation, parent feedback, and student achievement—in addition to the teacher-constructed portfolio. Using multiple measures also helps to allay the concern voiced by some administrators and teachers that a teacher might be an excellent portfolio maker but a poor teacher. We believe that this concern is exaggerated, but the perception itself is enough to undermine the credibility of the process. Hence, to offset the possibility of an ineffective teacher's receiving a high rating, as well as to strengthen the credibility of the portfolio process, it is wise for administrators to draw on multiple sources of information beyond the portfolio in evaluating a teacher.

Although specific scores should probably not be assigned to portfolios, detailed feedback is critical. Ideally, feedback on a teacher's portfolio should be presented in both written and oral form. However, time constraints may limit the amount of energy that administrators can devote to the feedback process. It is important to remember that feedback from the administrator is not the only, or even the primary, source of information for the teacher about his or

Section I
  Résumé
  Letters of Recommendation
  Educational Philosophy

Section II
  Unit Plan Overview
  Lesson Plan
  Reflective Summary
  Artifacts

Section III
  Professional Development Plan
  Professional Development Expertise

Figure 8.2. Sample Portfolio Table of Contents

her practice. Regular portfolio-based conversations among teachers should be an integral and ongoing feature of the portfoiio process.

Examples From Teaching Portfolios

In this section, we present examples from two teachers' portfolios. The first examples come from the portfolio of Nancy Hall, a first-grade teacher in a preservice education program at the University of Denver in Colorado. Preservice teachers at the University of Denver are required to prepare portfolios as part of their professional preparation program. The second set of examples comes from the portfolio of Janet Junkin, an experienced middle school social studies teacher in the Douglas County School District in Colorado, who submitted this portfolio to the district's Outstanding Teacher Program. Teachers in the Douglas County School District are invited to present portfolios to qualify for a designation of "outstanding" and a $1,000 bonus for that school year.

An Elementary School Teacher's Portfolio

First-grade teacher Nancy Hall placed the table of contents shown in Figure 8.2 in her portfolio. The first section of her portfolio provides background information about her experiences and perspectives, the second section presents direct evidence of her teaching, and the third section describes her future professional development plans and evidence of her professional expertise in one area of education (cooperative learning).

In a one-page statement of her philosophy of education, she described her academic instruction in the following way:

> My approach to teaching will incorporate the integration of the four academic disciplines (math, social studies, science, and language arts), as well as the arts. For example, study of the tropical rain forest would include learning about the kinds of trees, animals, and weather patterns that are particular to those regions. The height of the jungle trees could be compared to the indigenous trees of Colorado. The word "tropical" would generate a geography lesson. There is [abundant] literature related to the rain forest that could lead to a discussion on social issues and ecology. Through art projects the classroom could be transformed into a multi-rain forest. Our daily lives are interconnected with things around us, and it makes sense that we approach education in the same manner.

Nancy Hall also incorporated several overviews of a unit that she taught on the rain forest. One overview was presented in the form of a calendar, with each day's activities briefly summarized. In another overview, she described each week's focus in greater detail. Here is her description of week 3's events:

> This week the class will learn about biodiversity and why it's important. We will begin the study of rain forest animals and feature a different animal each day. The student will learn about the characteristics, classification, habits, and habitat of each animal. This week we will study the toucan, hummingbird, sloth and armadillo. I will incorporate [information about] the animals in the math journal problems.

She also included a lesson plan from the unit, examples of student work, assessments from the students about the value of the unit, and the university supervisor's visitation report on the lesson documented in the portfolio.

Nancy Hall then offered a one-page reflective summary of her lesson on the monkey. She noted the following strengths and weaknesses in her instruction:

> I felt that the lesson on the monkey went particularly well. It was successful because of the following reasons: the information itself was of interest and I planned two activities which require student participation. To demonstrate the sensitivity of finger tips and the precision grip that primates use, I blindfolded [a student] and had him pick up and identify beans, nuts, and raisins. I chose a few students to pantomime phrases I had written down to prove that communication is achieved through means other than words and that monkeys have this capacity. I have found interactive lessons to be rewarding and plan to involve the students in demonstrations of various concepts whenever possible.
>
> I didn't feel that the student surveys I provided gave an accurate picture of what the students thought about the lesson. I would not repeat the survey assessment. I felt that a better evaluation of the lesson were the reactions and information generated during the "think, pair, share." As the students verbalized information it was clear that they had listened, learned, and enjoyed the lesson.

A Middle School Teacher's Portfolio

Middle school social studies teacher Janet Junkin prepared a portfolio for her district's Outstanding Teacher Program. In her portfolio, she explained why she believed her performance met the district's criteria for "outstanding" for each of the three content standards. In this excerpt, she addresses the content standard "collaboration and partnership":

> I believe that my performance is outstanding in the area of collaboration and partnership because I believe and model the philosophy, "It takes a whole village to raise a child." Quality education occurs when the school staff, parents, and community contribute to this process as a team. . . .
>
> I visit my core students in their elective classes throughout the year for a variety of reasons. I enjoy seeing "my kids" in another

> setting. I value our elective program and I want the students to know this. This provides an avenue for collaboration with other teachers to benefit my students. . . .
>
> Recruiting and working with parent volunteers through the magazine sale, geography bee, student council, team field trips, community service projects, and the cruise is a very positive experience for me and has accomplished many things to benefit our students. . . .
>
> I draw on the community and private sector through phone calls, letters, and personal contacts. . . . These efforts provide my students with additional classroom resources as well as support for school-wide events.

To support her claim of outstanding performance in the area of "content and pedagogy," for example, Janet Junkin included information associated with a unit that she taught on Western Europe. In one activity, she asked the students to prepare an orientation brochure for American families who were about to move to Western Europe. Instructions that she gave to the students, which were documented in her portfolio, included the following:

> You work for a major American corporation that is expanding its business overseas . . . to Western Europe. You have been promoted to the position of Director of Expatriate Orientation. You are to provide training and information for FAMILIES about to move overseas.

She asked the students to begin the assignment by first identifying the questions that they might address in the orientation brochure:

- Three questions YOU might ask if you found out your family was moving overseas
- Three questions your SIBLING might want answered
- Three questions one of your PARENTS might want answered

As evidence of what students had learned, she included one student's brochure, called "A Traveler's Guide to France," in her portfolio. The brochure gave information about topics such as attractions, climate, and holidays.

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Observed
1. This teacher believes that ALL students can learn.	[✓ SA]	A	D	SD	NO
Comments: *She offers extra opportunities and experiences for studemts such as the Geography Bee.*					
2. This teacher makes adjustments for individual differences among students.	[✓ SA]	A	D	SD	NO
Comments: *They can choose what they want to do on projects.*					
3. This teacher uses a variety of strategies to meet my child's needs.	[✓ SA]	A	D	SD	NO
Comments: *There are creative projects that help them learn.*					

Figure 8.3. Client Survey, Douglas County School District

In addition to a variety of work samples from herself and her students, Janet Junkin also included surveys from students and their parents about her effectiveness as a teacher. These surveys, which were required for all teachers participating in the Outstanding Teacher Program, were mailed to 20 different families. The surveys were designed to parallel the content standards for the district's teachers. Excerpts from one family's responses to the 13-question survey are presented in Figure 8.3.

These surveys were anonymously returned directly to the teachers, who then included commentary in their portfolios about what they learned from the surveys. In her one-page reflections on the client surveys, Janet Junkin stated that she learned the following:

> My clients value guest speakers and a focus on current events. Biweekly progress reports are very time-consuming, but based on these surveys they are a very successful form of communication with my clients.

After she turned in her portfolio to her administrator, Janet Junkin received an overall rating as well as written feedback from her

Content Standard: Content and Pedagogy	
Areas of Strength	*Areas for Growth*
Uses wide variety of techniques. Connects to outside experiences. Celebrates diversity. Plans interdisciplinary units. Employs authentic assessment. Piloted district standards.	If possible, carry out interdisciplinary units with core teammates as well as with elective teachers.

Figure 8.4. Administrator Feedback on Janet Junkin's Portfolio

administrator about aspects of her performance for each of the district's three content standards (Figure 8.4 presents an example for one of the standards). Based on her portfolio, Janet Junkin was awarded a designation of "outstanding."

The examples presented above from the portfolios of Nancy Hall and Janet Junkin are intended to suggest the possibilities for what might be included in a portfolio. Ultimately, the contents that are selected will vary, depending on the purposes of the portfolio, expectations for teacher performance, school improvement plans, and individual teachers' interests and needs.

Putting Portfolios Into Practice

In the following pages we offer some suggestions for putting portfolios into practice. We propose a series of steps that administrators might follow in introducing portfolios in their schools or districts, as well as considerations for creating a productive climate for teachers who are willing to try out portfolios:

1. Define the expectations for teacher performance.
2. Clarify the purposes for the portfolio.
3. Identify the products for the portfolio.
4. Develop guidelines for portfolio construction.
5. Establish procedures for portfolio evaluation.

Step 1: Define the expectations for teacher performance in a building or district. Before portfolios are actually introduced, it is critical that administrators first examine the expectations, or content standards, for teachers in their buildings or districts. What are teachers expected to know and be able to do? Are the expectations for performance for beginners different from those for experienced teachers? Many good examples of content standards currently exist (see Table 8.1, for example), and these can be adapted for use in other settings.

Step 2: Clarify the purposes for the portfolio. Decisions about the purposes of the portfolio will influence issues related to the portfolio design, construction, and evaluation. If the primary use of the portfolio will be for evaluation, then the portfolio process should be similar for all participants; otherwise, administrators may not be able to evaluate the portfolios in a fair and consistent fashion. On the other hand, portfolios that are intended primarily for professional development may be completely different from teacher to teacher in format and focus. In this instance, the administrator's primary responsibility is not to evaluate the portfolio but to provide feedback to the portfolio's creator. Clarity about these purposes will help prevent the portfolio process from becoming a paper chase for teachers or an evaluation nightmare for administrators.

Step 3: Identify the products for the portfolio. Although many different products can be productively placed in a portfolio, requirements (or recommendations) for the inclusion of certain types of information will allow teachers to discuss their instruction with each other more easily and will enable administrators to evaluate portfolios more efficiently. What types of information should be included in portfolios? We have offered a general list above, but an even more detailed set of products and specifications is recommended. For example, a school or district might require that each teacher submit a one-page narrative overview of three consecutive lessons and a two-page lesson plan for the second lesson in the series. Again, it is worth mentioning that if the portfolio is not being used for evaluation, the need for consistency in the type and number of portfolio products is much less critical.

As the portfolio products are being identified, it is important to link the products to the expectations, or content standards, for teachers. The question to be answered, then, is, What kinds of information in a portfolio can best provide insights into a teacher's performance for a particular standard? A matrix is a useful tool for examining the relationship between the portfolio products and content standards (see Figure 8.5). For example, for what content standards

	Lesson Plan	*Student Assessments*	*Parent Feedback*
Assessment and instruction	X	X	
Content and pedagogy	X		X
Collaboration and partnership			X

Figure 8.5. Matrix of Content Standards by Portfolio Contents

might a lesson plan, student assessments, and parent feedback provide evidence?

Step 4: Develop guidelines for the portfolio construction process. It is important to establish guidelines for portfolio construction so that teachers will have to spend as little time as possible thinking about portfolio logistics and as much time as possible considering issues of teaching and learning. These guidelines should include such elements as the timeline for completing the portfolio, required or recommended contents, specifications for preparing information (e.g., date and caption each piece of information), and information about the evaluation process. This information might be presented in a handbook or in a series of one-page handouts.

Step 5: Establish procedures for evaluating the portfolio. In evaluating portfolio performance, a number of issues need to be addressed. Who will evaluate the portfolio—the principal, a trusted peer, a committee? How will the information be used—as part of an individual evaluation, for self-assessment, for discussion about school goals? What kind of feedback will a teacher receive—a score, written comments, a conference with a supervisor? Teacher confidence in the program will be greatly enhanced if evaluation-related decisions such as these are made well in advance.

Introducing the Use of Portfolios

Along with those described above, the following steps should be followed during the introduction of the use of portfolios in a school or district:

- Enlist volunteers.
- Start small.
- Keep the risk low.
- Encourage portfolio-based conversations.
- Use multiple measures of evaluation.

Enlist volunteers. A core group of enthusiastic volunteers should be identified who are willing to be carpenters constructing their own portfolios at the same time they are architects building the overall program. These teachers, then, will not only be trying out the process by preparing their own portfolios, they will also take leadership roles in tasks such as identifying the content standards and developing portfolio procedures. Many new ideas in education fail not because the concepts are flawed, but because they did not have teacher input or buy-in.

Start small. Begin by having teachers focus on only a single content standard or school goal, or only on a few types of portfolio products. Portfolios can quickly become unmanageable if a lot of information is collected without clear purposes and procedures for their organization and use. After the initial design issues have been worked out, and portfolios have been successfully introduced, then the focus of portfolios can be expanded to more content standards or multiple goals or a greater variety of information.

Keep the risk low. The portfolio pioneers should feel confident that the work that they do will, at a minimum, have no negative consequences. Clarity about the agenda of the program and the uses for the portfolios will help to establish an atmosphere of trust. Moreover, not only should the risk remain low, but providing incentives for teachers, such as release days or bonuses, will increase the likelihood of the program's taking root.

Encourage portfolio-based conversations. Portfolios provide an opportunity for teachers to work together to foster their professional development and to address difficult problems of practice. Administrators should promote these collaborative interactions through a variety of strategies, such as devoting a portion of faculty meetings to portfolio-based conversations or structuring schedules so that teachers have shared planning times. Wolf, Hagerty, and Whinery (1995) offer suggestions for organizing and conducting portfolio-based conversations among teachers.

Use multiple measures of evaluation. Like any assessment approach, portfolios have strengths as well as limitations. They may provide a clear view of a teacher's instructional strategies, for example, but give no information about that teacher's classroom management practices. Furthermore, portfolios are better at revealing strengths than they are at providing information about weaknesses. For these reasons, portfolios should be part of an overall evaluation system that might include classroom observation, evidence of student achievement, and client feedback.

Conclusion

In this chapter, we have defined the concept of a teaching portfolio and discussed the various forms that portfolios can take. We have also provided examples from actual portfolios. Finally, we have suggested steps that administrators might take in putting portfolios into practice.

In essence, we believe that all portfolios should contain carefully selected examples of teacher and student work, framed by commentaries and captions, and brought to life through extended conversations with colleagues and supervisors. We also emphasize that a portfolio's purpose drives many decisions about the specific contents of the portfolio, as well as the process for constructing and evaluating it.

Whether teachers are creating portfolios as part of a school evaluation plan or for their own professional development purposes, we believe that the process of carefully documenting and reflecting on selected aspects of one's practice enhances performance. Thus, although the purposes for creating portfolios may vary, all portfolios contribute to the same ultimate aim—to advance student learning through the professional development of teachers.

References

Borko, H., Michaelec, P., & Timmons, M. (1996). *Student teaching portfolios: A tool for promoting reflective practice.* Manuscript submitted for publication.

Lichtenstein, G. (1997). *Portfolios for hire.* Unpublished manuscript.

Lichtenstein, G., Wolf, K., Pease, P., Ruckle, C., & Campbell, G. (1996). *Teaching portfolios and relicensure in Colorado.* Manuscript submitted for publication.

National Board for Professional Teaching Standards. (1989). *Toward high and rigorous standards for the teaching profession.* Washington, DC: Author.

Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*(3), 36-41. (ERIC Document Reproduction Service No. EJ 385 344)

Shulman, L. (1992). *Portfolios for teacher education: A component of reflective teacher education.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Stronge, J. H., & Helm, V. M. (1991). *Evaluating professional support personnel in education.* Newbury Park, CA: Sage.

Wolf, K. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan, 73,* 129-136.

Wolf, K. (1996). Developing an effective teaching portfolio. *Educational Leadership, 53*(6), 34-37.

Wolf, K., Hagerty, P., & Whinery, B. (1995). Teaching portfolios and portfolio conversations for teachers and teacher educators. *Action in Teacher Education, 17,* 30-39.

Wolf, K., Lichtenstein, G., Bartlett, E., & Hartman, D. (1996). Professional development and teaching portfolios: The Douglas County School District Outstanding Teacher Program. *Journal of Personnel Evaluation in Education, 10,* 279-286.

9

Teacher Self-Evaluation

PETER W. AIRASIAN
ARLEN GULLICKSON

Teacher self-evaluation is a process in which teachers make judgments about the adequacy and effectiveness of their own knowledge, performance, beliefs, and effects for the purpose of self-improvement. At the heart of teacher self-evaluation is the belief that teachers' main reasons for engaging in professional development activities come from their own experiences of what it means to be a teacher and from a personal need to understand, critique, and improve their own practice (Airasian & Gullickson, 1994a, 1994b; Hargreaves, 1995). In self-evaluation, it is the teacher who collects, interprets, and judges information bearing on personal practice. It is the teacher who frames criteria and standards to judge the adequacy of his or her beliefs, knowledge, skills, and effectiveness. It is the teacher who decides on the nature of professional development activities to be undertaken.

Self-evaluation is related to such concepts as the reflective practitioner, reflection on practice, teacher connoisseurship and criticism, teacher research, self-understanding, educational action research, and analysis of practice. All of these approaches encourage teachers to assess their own practice, based on the belief that if meaningful professional development is to occur, the focus must be in areas where teachers perceive the need for improvement (Sergiovanni & Starratt, 1988). Self-evaluation has been likened to a formative evaluation of one's practice, formative in the sense that the evaluation is aimed at examining and, if necessary, altering or improving practice (Airasian & Gullickson, 1994a; Barber, 1990; Kremer-Hayon, 1993). Some have viewed it as an informal type of action research based in and concerned with practice (Gitlin et al., 1992; Hatton & Smith, 1995; Osterman & Kottkamp, 1993).

Underlying teacher self-evaluation are the following beliefs (Airasian & Gullickson, 1994b; Osterman & Kottkamp, 1993):

- Teachers need professional growth opportunities.
- Teachers want to improve their practice and knowledge; teachers want and need information about their knowledge, performance, and effectiveness.
- Teachers are capable of assuming responsibility for much of their own professional growth and development, given time, encouragement, and resources.
- Collaboration enriches professional growth and development.

It is, however, necessary to point out that in spite of the importance and usefulness of teacher self-evaluation, it is only part of a viable, ongoing program of teacher evaluation. Self-evaluation is evaluation for teachers to help them become aware of, reflect on, analyze, and make decisions about their practice. However, in most school settings there is a need for other types of teacher evaluation, because teachers are accountable to many constituents beside themselves: (a) pupils and the community who are affected by their actions, (b) employers with whom they have contractual obligations, and (c) their profession, which has its own standards and expectations (Dwyer & Stufflebeam, 1996). Although self-evaluation can fulfill some of these accountability needs, it cannot fulfill them all; thus, other forms of teacher evaluation are necessary.

Why Is Teacher Self-Evaluation Important?

The nature of teacher evaluation is changing in response to three trends: poor pupil achievement, bottom-up reform efforts, and new views of teaching and learning. First, the perceived poor performance of American students on national and international tests has been associated, in part, with unsatisfactory performance by teachers. Critics argue that improving the performance and quality of teachers is a crucial and critical prerequisite of improving student achievement (Carnegie Task Force on Teaching as a Profession, 1986). As a consequence, programs emphasizing the importance of careful preservice and in-service teacher evaluation, accompanied by raised standards for teacher certification, have been implemented in many states and school districts (Darling-Hammond, 1990; Dwyer & Stufflebeam, 1996; Grimmett, Rostad, & Ford, 1992).

Second, a current emphasis in teacher evaluation is bottom-up approaches to educational reform (Glickman, 1992; Hargreaves & Fullan, 1992; Smith & O'Day, 1991). In contrast to the top-down reform approaches of the 1970s and 1980s (Milner, 1991), which were based largely on mandated, common evaluation strategies for all teachers in a state or district, bottom-up reforms are focused on the individual school or teacher and seek to make teaching more professional by providing varied and ongoing opportunities for professional development. There is a more balanced emphasis on summative and formative evaluations. This shift can be seen in efforts to do the following:

- Extend assistance to new teachers
- Broaden career opportunities for experienced teachers
- Make teachers more responsible for demonstrating their own competence
- Introduce portfolio evaluation to foster teacher reflection
- Provide opportunities for collegial and cooperative teacher activities, such as site-based management and action research

All these efforts are designed to help teachers play an informed and active role in assessing and advancing their own practice.

The bottom-up approach views teachers as influential educational decision makers and seeks to use evaluation to identify the professional development they need to enhance their own practice (Airasian, 1993). There is recognition that experience is a necessary but not sufficient prerequisite of teaching expertise. What distinguishes the expert teacher from the nonexpert is the former's ability to learn from reflection and experience (Berliner, 1987; Schön, 1987).

Evaluation, including self-evaluation, is viewed as a major and important means of making teachers aware of their practice, challenging them to think about it, and encouraging them to analyze and evaluate it and implement changes as needed.

Third, the increased emphasis on teacher evaluation occurs in response to the growing attention accorded the teaching and learning of higher-level thinking skills. The view of teaching and learning one adopts has direct implications for what and how one teaches (Delandshire & Petrosky, 1994; Driver, Asoko, Leach, Mortimer, & Scott, 1994), and consequently for how one's teaching should be evaluated. A view of learning that conceives of individuals (including teachers) as constructors of their own knowledge (Bruner, 1990; Wertsch & Toma, 1995) rather than as passive recipients of others' knowledge calls for supervision and teacher evaluation models different from those currently in common use. The search is for models that can help teachers acquire a deeper personal understanding of the teaching-learning process and a reconstructed role perception of themselves as active promoters of that process (Carter & Doyle, 1987; Schön, 1988). There is a growing balance between an external, inspectorial view of teacher evaluation and one in which responsibility is given to teachers to self-evaluate.

Stufflebeam and Shinkfield (1985) suggest that the most important evaluations of professionals are the ones conducted by the professionals themselves. Opportunities to explore their beliefs, knowledge, practices, and effects make teachers more likely to question their taken-for-granted expectations, norms, beliefs, and practices. Clandinin and Connelly (1988) emphasize that self-knowledge and self-understanding are the keys to professional growth, and that the best way to improve these is to give teachers more control and responsibility for self-evaluation. Gitlin et al. (1992) argue that building upon notions such as the self-monitoring teacher and the teacher as classroom researcher represents the best way to give teachers a true "voice" in their practice. Kuhn (1991) emphasizes the importance of thinking about one's theories and beliefs, not just applying them. Tom (1984) describes an effective teacher as one who "is able to conceive of his teaching in purposeful terms, analyze a particular teaching problem, attempt the approach, judge the results in relation to the original purpose, and reconsider either the teaching or the original purpose" (p. 94). Gadamer (1975) notes that reflection on practice is a powerful method for bringing one's tacit, unquestioned beliefs and assumptions into the open to be self-examined. Although none of these commentators uses the term *self-evaluation* to describe the processes recommended, each recognizes the centrality of teachers' self-evaluation activities to their development and professional growth.

In sum, teacher self-evaluation is an important process for teachers to utilize for all of the following reasons (Airasian & Gullickson, 1995, 1996; McColskey & Egelson, 1993):

1. It is a professional responsibility to do so.
2. It focuses professional development and improvement at the classroom or school level, where teachers have their greatest expertise and impact. It grows from the immediacy and complexity of the classroom, as do teachers' motives and incentives.
3. It recognizes that organizational change is usually the result of individuals' changing themselves and their personal practice, not of top-down mandates.
4. It gives teachers "voice," that is, a stake in and control over their own practice.
5. It makes teachers aware of the strengths and weaknesses of their practice.
6. It encourages ongoing teacher development and discourages unchanging classroom beliefs, routines, and methods.
7. It treats teachers as professionals and can improve teacher morale and motivation.
8. It encourages collegial interactions and discussions about teaching.

What Is the Role of Reflection in Teacher Self-Evaluation?

Integral to teacher self-evaluation is reflection, the act of thinking about or pondering an idea, issue, or problem. Schön (1983, 1987) and others (Calderhead, 1988; Kuhn, 1991; Osterman & Kottkamp, 1993) have noted that teachers must reflect on their practice in order to understand, critique, and modify it; reflection is a central aspect of the process of constructing knowledge and developing professionally.

Accountability to oneself is a fundamental aspect of professionalism because it implies a moral commitment to serve the interests of students and an obligation to review (i.e., reflect about) the nature of one's own practice in order to improve it (Eraut, 1993).

Recently, the work of Schön (1983, 1987, 1988) has been the focus of discussion about teacher reflection. Although we choose to focus on the work of Schön, we note that other, more detailed models of teacher reflection have been advanced (Hatton & Smith, 1995; Louden, 1991, 1992; Valli, 1992; Van Manen, 1991). These models extend—and to some extent complicate—Schön's conceptualization of reflection and appear to be more useful for studying the teaching *process* per se than teacher self-evaluation.

Schön proposes two ways in which teachers reflect on their practice: *reflection in action* and *reflection on action*. The former focuses on the spontaneous and tacit reflections and decisions teachers make when they are engaged in the process of teaching. During reflection in action, the teacher reflects on an activity that is in progress, lets the activity, in Schön's words, "talk back to" him or her, and makes an on-the-spot decision to alter or maintain the activity. The process is largely tacit and spontaneous.

Reflection on action takes place out of the activity of practice; it is consideration of an action, belief, or effect divorced in time from the factors that prompted the need for reflection. Reflection on action, with its temporal separation of activity and reflection, comes closer to the commonly held understanding of *reflection* than does reflection in action (Airasian & Gullickson, 1994b; Osterman & Kottkamp, 1993). The process is more conscious and reasoned. Note that both reflection in and reflection on action can produce teacher decisions, although the nature and consciousness of the decision-making process differ substantially for each. However, given the differences between the two—the former being tacit and spontaneous, deriving self-knowledge from in-practice reflection and the latter being conscious and planned, deriving self-knowledge from out-of-practice reflection—it is not surprising that they differ in many ways.

Reflection in Action

Given classroom complexity and uncertainty, teachers must learn much of their teaching repertoire from the concrete reality of their classroom practice—from the specific problems they must respond to and manage (Elbaz, 1990; Ross, Cornett, & McCutcheon, 1992). From trial-and-error, self-evaluative experiences, teachers forge a "personal practical knowledge" that structures their teaching activities (Kleinsasser, 1992; Schwab, 1969). The world of the classroom demands that teachers have this practical knowledge to enable them to negotiate the practical problems they face.

Without experientially based personal knowledge, practical activities such as teaching cannot be fully understood or practiced. Experience creates problems for the practitioner and the necessity of finding solutions. In this scenario, the role of tacit/spontaneous reflection in action appears crucial and inevitable. The informal, experiential, trial-and-error nature of teachers' in-action activities is an important self-evaluation medium through which teachers become learners and constructors of their own practical knowledge and beliefs. The nature of classrooms leads teachers to initiate self- evaluation. "After all, it's the teacher's subjective school-related knowledge which determines for the most part what happens in the classroom; whether the teacher can articulate his/her knowledge or not" (Halkes & Olson, 1984, p. 1).

Although teachers' tacit, in-action reflections are important in shaping their practice and teaching beliefs, this approach has many characteristics that hamper its usefulness as a self-evaluation strategy for professional growth and development. We noted a number of these characteristics previously: The tacit, fleeting nature of reflection in action makes it difficult to study and model; the amount of time available to reflect on any particular action event is very small, leading to brief, often shallow, reflections and decisions; the fleeting nature of reflection in action can result in superficial, expedient decisions that lead to simplistic explanations and heuristics (Court, 1988; Elbaz, 1990; Farnham-Diggory, 1994). Additionally, the bases for reflection in action are particular to each teacher, thereby encouraging the formation of idiosyncratic practice (Gilliss, 1988). In the extreme, idiosyncrasy can provide a justification for teachers to dismiss or fail to seek out information that might lead to improved practice.

The fact that reflection in action is a personal, spontaneous, tacit process means that it limits teachers' opportunity to seek or receive feedback from others or from the reflective episode itself (Kremer-

Hayon, 1993). Only the teacher's fleeting impressions and "gut-level" instincts are available for him or her to identify weaknesses and make decisions about practice. Moreover, reflections in action focus on events and situations that arise spontaneously and unplanned (Cochran-Smith & Lytle, 1990; Kagan, 1992). In a real sense, what is reflected upon and evaluated in action is controlled not by the teacher, but by the spontaneous flow of events transpiring in the classroom. Thus, reflection in action is reactive, not proactive. There is no opportunity to identify procedures, alternatives, or options before action takes place. This reactivity serves to narrow one's view of practice. Finally, reflection in action, like all other self-evaluation approaches, shares an inherent and problematic characteristic: The teacher is both evaluator and evaluatee (Airasian & Gullickson, 1994b; Kremer-Hayon, 1993). Thus, even if a teacher overcomes the many limitations of reflection in action described above, he or she still must face the difficult barrier of personal biases in decision making.

Reflection on Action

Reflection on action involves a post hoc, conscious replay of an activity or event from practice. It is a conscious rather than a tacit process, and it is controlled by the teacher, not spontaneously occurring. The ability to revisit and linger over areas or events has many potential benefits for teachers. First, it permits conscious reflection related to practice; thus, it is accompanied by a degree of awareness of actual practice not usually available from reflection in action. Second, the time devoted to reflection on action can provide a deep, broad perspective on practice. Third, conscious reflection on classroom action and events plays nicely into the cyclical nature of teaching. Inevitably, teaching involves repeating many processes (e.g., lesson planning, grading, cooperative learning) and content topics (e.g., distinguishing use of *lay* and *lie*, rate/distance/time problems, balancing chemical equations) over the course of a week, month, or school year. Reflection on action can be a powerful tool for identifying, anticipating, and making changes in cyclical processes and content areas. In many cases, the alternatives identified by reflection on action at one point in time can be recalled and applied to the same process or topic at a subsequent point in time. It is for this reason many teacher education programs have begun to emphasize and develop experiences designed to spur preservice teachers' reflections on their classroom actions (e.g., Gore & Zeichner, 1991; Hatton & Smith, 1995; Valli, 1992).

However, like reflection in action, reflection on action has its limitations as a self-evaluation strategy. First and foremost, conscious reflection is not often associated with teaching, and teachers are not generally provided with instruction on how to be effective reflective practitioners (Elbaz, 1990; McNamara, 1990). As a consequence, teachers may consider the wrong information, weigh it improperly, ask the wrong questions of it, and make the wrong inferences when using it (Nisbett & Ross, 1980). These problems are accentuated by teachers' tendency to adopt simple, often superficial explanations for classroom events (Feiman-Nemser & Floden, 1986; Peterson & Barger, 1985).

Second, in order for reflection to lead to self-evaluation, it must result in a teacher decision. For many reasons, reflection may stop at the revisiting stage, without a decision or strategy being made or identified. Teachers are no different from other people who reflect on their activities and actions, insofar as a substantial number of both the reflections and decisions made from the reflections are forgotten from one day to the next. Self-evaluation resulting from reflection on action is highly dependent on the teacher's recall of things reflected on.

Third, reflection on action is practice driven. Even with conscious deliberation, reflection on action takes place in a closed system in which the teacher is both evaluator and evaluatee. Teachers' preexisting beliefs and experiences are a strong influence on the interpretive frameworks they use to label and categorize new reflective situations (Shulman, 1987). Frequently, such preconceptions preempt adequate teacher reflection on situations of concern (Wildman & Niles, 1987).

Because of the personal nature of reflection and self-evaluation, interpretation and decision making inevitably pit reason and objectivity against emotion and self-protection. This issue has been raised in a variety of ways by a number of commentators (Barber, 1990; Kuhn, 1991). Knowledge about teacher self-evaluation is not voluminous, but the studies noted above do show differences between teachers' own reflections, awareness, and self-evaluations and those made by the teachers' supervisors, peers, and students. Such findings raise questions about teachers' objectivity when it comes to judging their own practice. Another important concern is that incompetent or inexperienced teachers, regardless of the amount of conscious reflection they engage in, will produce reflections that prevent them from recognizing that they are performing at an unsatisfactory level or will be unable to identify means to improve even when they know their performance is unsatisfactory.

Thus, although reflection on action does provide a more deliberative, conscious examination of practice than that provided by reflection in action, it is clear that reflection of either type is laden with problems that inhibit its leading to satisfactory self-evaluation. It is important to point out that many of these problems are inevitable, because self-evaluation means that the teacher personally must be the reflector, interpreter, and decision maker regarding practice. So, limitations such as subjectivity, personal interpretations, and preexisting perceptions and beliefs will always be issues of concern in teacher reflection and self-evaluation. However, there is an approach that has promise for diminishing these and other problems of teacher reflection, an approach based on obtaining externally generated information about practice.

Enhanced Reflection

The availability of external information can enhance teachers' reflection by providing them with explicit information about practice (Airasian & Gullickson, 1995; Osterman & Kottkamp, 1993; Shulman, 1987). As we noted previously, the reflective process that occurs during both reflection in and on action engenders many factors that inhibit a teacher's ability to recognize problems, delineate their salient features, marshal appropriate evidence for their interpretation, and make informed, meaningful decisions. One way to improve reflection and awareness of practice is to supplement teachers' personal perceptions and recall with formal evidence about practice.

Of course, efforts to supplement teachers' reflection about their practice with more formal, "objective" information are not new. An early emphasis on teacher reflection, awareness of practice, and self-evaluation is found in Flanders's work on interaction analysis in the 1960s (see Bailey, 1981; Travers, 1983). This work gave credence and respectability to the systematic study of teachers' behavior and also emphasized the need for teachers to be able to operate independently in classrooms, where there is little opportunity for external feedback on their own teaching behaviors.

The model of stimulating teacher self-evaluation through an outside observer or recorder remains predominant today, especially in evaluations of teacher practice and development. The function of the external observer or recorder is to help the teacher see himself or herself as others see him or her. The intent is to "confront" teachers with external information about their practice so they will see it, reflect on it, compare it with their own personal perceptions of practice, and evaluate it in a more objective light than when the only information about practice is self-generated (Fuller & Manning, 1973).

The intent exemplified in this approach to teacher self-evaluation underlies other common teacher evaluation approaches, such as microteaching (Allen & Ryan, 1969), interpersonal process recall (Kagan, 1975), self-report (Bailey, 1981), self-study materials (Silver & Hanson, 1980), and observations of teachers by school principals or department heads. All of these approaches are intended to enhance teacher awareness, reflection, and, subsequently, self-evaluation with formal, externally generated, information about practice.

However, the most common approach—principal observation and evaluation of teaching—is not regarded highly by most teachers as a basis for their self-evaluation. The infrequency and brevity of the typical observation, the observer's lack of knowledge of the classroom culture, the lack of specific suggestions for improvement, and the lack of support for improvement efforts are all cited by teachers as reasons principals' observations and evaluations rarely stimulate self-evaluation (Duke & Stiggins, 1985; Jones & Airasian, in press; Kottkamp, Provenzo, & Cohn, 1986).

A number of authors have compiled formal strategies that can be used to enhance teachers' awareness of practice and reflection. Angelo and Cross (1993) provide numerous examples of formal assessment strategies that can be employed by teachers to learn more about their practice. Kremer-Hayon (1993) describes the context of teacher self-evaluation, the relationship between pedagogical knowledge and self-evaluation, qualitative and quantitative methods for conducting teacher self-evaluation, and sources of support needed for self-evaluation in schools to be viable. Pollard and Tann

(1993) suggest a large number of reflective and self-evaluation activities aimed mainly at elementary school teachers. Saphier and Gower (1987) provide a thorough description of the knowledge base of teaching, with suggestions for areas in which to conduct self-evaluation. Haysom (1985) presents a number of self-evaluation strategies that relate to professional development. What is common among these and other sources (Airasian & Gullickson, 1996; Barber, 1987; Iwanicki & McEachern, 1983; McColskey & Egelson, 1993; Osterman & Kottkamp, 1993) is a reliance upon formal information gathering to support and enhance reflection in the hope that it will stimulate teacher awareness, recall, reflection, self-evaluation, and improvement.

Eight general strategies can be used to supplement and enhance teacher reflection, awareness of practice, and self-evaluation (Airasian & Gullickson, 1995):

- *Self-reflection tools:* These are designed to be completed by the teacher and allow him or her to evaluate performance in identified areas of practice, beliefs, and outcomes. Such tools usually take the form of checklists, rating scales, or questionnaires.
- *Media recording and analysis:* Media such as audio- and videotape may be used to record a sample of teacher performance for subsequent analysis by the teacher and/or peers. Microteaching is an example of media recording.
- *Student feedback:* This is gathered using instruments similar to self-reflection tools, except that pupils complete the forms instead of the teacher. Questionnaires, "minute surveys," and journals are examples of sources of student feedback.
- *Teacher portfolios:* Teachers prepare portfolios to provide evidence about their beliefs, knowledge, skills, and effectiveness. The process of collecting and analyzing portfolio pieces provides opportunity for self-evaluation. Note that a teacher portfolio can contain many of the self-evaluation strategies in this list.
- *Student performance data:* These include all student productions that can be used to help the teacher assess his or her own instructional effectiveness. Test results (classroom or standardized), essays, and projects are examples of student performance data.

- *External peer observation:* This involves a peer, colleague, or administrator observing and describing particular aspects of another teacher's behavior, such as questioning behavior, lesson organization, and student feedback.
- *Journaling:* This requires the teacher to maintain and reflect upon a record of classroom events or activities with the intent of identifying recurring problems, successes, and needs.
- *Collegial dialogue/experience sharing/joint problem solving:* This encourages collaboration and interchange among teachers around common problems. Procedures, strategies, and perceptions are shared. Exposure to the ideas and practices of colleagues is a potent strategy for examining one's own practice.

Use of these external strategies for gathering information about practice enhances teacher reflection and self-evaluation in a number of ways. First, external information about practice takes the teacher out of the closed circle of reflection in and reflection on action (Iwanicki & McEachern, 1983; Kremer-Hayon, 1993). It allows teachers to be proactive, not just reactive, when examining and evaluating their practice. Second, the availability of formal evidence-gathering tools allows the teacher to focus on areas beyond immediate practice. For example, self-reflection tools are available that allow a teacher to examine basic teaching beliefs, knowledge of recent innovations, and instructional processes (Airasian & Gullickson, 1996; Angelo & Cross, 1993). Microteaching can identify many aspects of practice that might not be questioned or identified as problematic in the context of direct practice.

Third, and quite important, the availability of formal, external information about an area of practice provides a comparative basis against which a teacher can examine his or her own perceptions and interpretations: Is the teacher's personal view corroborated by external evidence garnered from peers, colleagues, students, media, or some other source? Without this comparative base, the teacher might overlook needed areas of improvement or misjudge areas of strength (Airasian & Gullickson, 1994b; Barber, 1990; Osterman & Kottkamp, 1993). When a teacher's self-defined standards and criteria can be compared with externally generated information in the form of student feedback, peer observation, self-reflection instruments, media recording, and the like, the teacher has an external, independent basis against which to compare his or her personal reflections and perceptions. Note that this comparison cannot be made without some form of external information. Note also that the comparison is most powerful when the teacher makes his or her own predictions about what the external information will show *before* that information is actually obtained. The potential of an externally based description of practice to enhance and corroborate teachers' personal awareness of practice is an important argument for using formal, external, "objective" awareness of practice as an extension of reflection and an integral part of teacher self-evaluation.

Getting Started With Self-Evaluation

What is needed to carry out teacher self-evaluation for professional development? First and foremost, the impetus for involvement with self-evaluation must come from the teacher. Forcing teachers to self-evaluate destroys its purpose and success. However, there are many environmental conditions that affect teachers' willingness to participate in self-evaluation. It is these conditions that principals can manipulate to improve the climate and, thus, the likelihood that teachers will feel sufficiently comfortable and unthreatened to engage in self-evaluation.

Most of what has been written about teacher self-evaluation focuses on the teacher. This is natural because, as the name implies, teacher self-evaluation is a teacher-conducted activity. When peers, supervisors, principals, and students are mentioned in the context of teacher self-evaluation, it is usually as external reporters or observers who can provide objective information about teacher practice. We argue that such external information can foster and assist teacher self-evaluation, guide its use, and help sustain it when barriers threaten its use. Although external assistance can be provided by many persons, in most schools it is the principal who has direct authority and responsibility for providing motivation, guidance, and support for teacher activities.

As we noted previously, teachers regularly employ reflection in action during their teaching, less regularly employ reflection on action, and rarely engage in enhanced reflection on action. The challenge to principals is to increase the strongest of these three modes, enhanced teacher self-evaluation, to serve the development of teachers. We believe school principals are the best persons to help teachers address this challenge, by making sure that both they and their teachers (a) know what self-evaluation is and know some self-evaluation strategies teachers can use, (b) know the characteristics teachers must have to do viable self-evaluation, and (c) construct a school environment conducive to teacher self-evaluation. We turn now to strategies of self-evaluation and the teacher characteristics and school support structures needed for successful teacher self-evaluation.

Strategies for Teacher Self-Evaluation

The preceding sections have provided common understandings about what teacher self-evaluation is and why it is important. In this and succeeding sections, we address the issue of what a principal can do to get self-evaluation started. We present a few specific examples of self-evaluation strategies that can be shared with teachers to provide understanding of the nature and intent of self-evaluation. The examples are meant to be illustrative of possibilities, not exhaustive and all-encompassing. The examples are excerpted from our *Teacher Self-Evaluation Tool Kit* (Airasian & Gullickson, 1996); interested readers are referred to this publication for additional examples.

Two fundamental criteria underlie the examples to be presented. First, the examples are formal, in the sense that they produce some tangible record or information for the teacher to use for self-evaluation. Although the nature and source of this information can vary considerably, the aim is to produce tangible information to use in analyzing and making decisions about practice. It is important to note, however, that the focus on formal approaches does not mean that the approaches have to be complicated, time-consuming, or controlled by others.

Second, the examples are aimed at helping teachers to obtain an awareness of their practice, including their beliefs, knowledge, activities, and effects. Much of the current discussion about teacher development emphasizes teacher reflection and introspection as the path to change and improvement. Although reflection and introspection are important aspects of professional development, they are insufficient to guarantee improvement. The missing link between reflection/introspection and the self-evaluation decisions that lead to professional development is an objective awareness of one's practice. Before we can meaningfully reflect on practice, we must have a clear awareness of that practice. Awareness is essential for change, and although reflection may help awareness, it does so much more productively when it is based on formal information about practice.

The following examples vary in many ways, but there are some common guidelines that apply to any self-evaluation approach: anonymity, focus, comparison, and multiple tools.

1. Self-evaluation strategies that obtain feedback from students should protect the anonymity of the student respondents. Maintenance of student anonymity protects them and encourages them to provide the honest feedback needed for viable self-evaluation.
2. In selecting areas to self-assess, teachers should focus on narrow aspects of teaching rather than more general ones. It is more productive and diagnostic to examine one's "questioning strategies" or "beliefs about learning" than it is to inquire about one's general "teaching competence" or "instructional approaches."
3. Whenever possible, it is advantageous to build a comparison into self-evaluations. Before administering a self-evaluation tool, teachers should attempt to predict what the main outcomes will be. If possible, these predictions should be made in writing. Then, when teachers get the actual results, they can compare them with their predictions.
4. No single strategy can provide all the information needed to self-assess teaching. Different self-evaluation strategies are useful for evaluating different areas, as is shown in the following examples.

Example 1: Examining Teacher Practice or Effects Using a Student Feedback Tool

The minute survey (Angelo & Cross, 1993) is a very flexible and simple strategy for getting feedback from pupils regarding various aspects of classroom activities. To conduct a minute survey, the teacher sets aside 2 or 3 minutes at the end of a class period to survey the students about some aspect of the lesson. The teacher then asks the students to answer one or two questions about the lesson (either by writing down their answers or, in early grades, raising their hands or making some other indication). Students respond anonymously. For example, the teacher might ask the students to address one or two of the following issues:

1. What are two things you learned from the lesson?
2. Write one question you would like to have answered about the lesson.
3. What part of the lesson would you like to know more about?
4. How confident are you about doing tonight's homework on this lesson?
5. Rate the success of the example presented to reinforce the lesson's main point.
6. Rate how well the reading assignment prepared you for today's lesson.
7. Solve one or two problems or calculations similar to those taught in the lesson.

There are, of course, many other questions a teacher could ask pupils about a lesson, an assignment, a homework exercise, a field trip, a video presentation, a demonstration, or even a test. However, to make the use of the minute survey most informative, the teacher should follow these guidelines:

1. Let students reply anonymously.
2. Keep the amount of writing requested of the students small. Don't ask students to write responses to four or five questions. The minute survey is meant to be completed in a short time; hence, it should focus on only one or two aspects of the lesson.
3. Before reading the responses, try to answer the questions yourself, based on your perception of the lesson. For example, what do you (the teacher) think will be the two most important things most students will say they learned, or the one question most students will want answered, or the degree to which the example you showed helped students to learn? Answering the questions yourself will provide a check on your observational impressions in light of students' actual responses.

4. Read the responses, compare them with your own predictions, and, depending on the responses, use the information to reteach, answer questions, reexplain, or move on to the next topic.

Example 2: Examining Practice Using Media, an External Observer, or Student Feedback

There are many ways for a teacher to obtain information about the nature and content of classroom discourse. He or she can have a colleague observe a lesson or two and provide information about who talks most, about what, the quality of questions asked, the reinforcement and encouragement given students, and so on. The teacher can also poll pupils to obtain their impressions of the nature and content of classroom discourse, or use media such as audio- or videotape to record and subsequently analyze discourse. One advantage of taping classroom interaction is that it provides a permanent record of performance that can be examined and reexamined. Often, when teachers are engaged and engrossed in the process of teaching, they are not aware of how much they talk, to whom, why, and with what purpose. This strategy also can be effective if a teacher collaborates with another person over time (McColskey & Egelson, 1993).

Regardless of the method used to obtain information on classroom discourse, it is useful for the teacher to do two things to increase the usefulness of the information:

1. State a set of questions he or she wants to answer before collecting information.
2. Answer the questions himself or herself before examining the information collected.

Sample questions that might guide analysis of discourse during a lesson include the following:

1. What proportion of time is spent in teacher talk?
2. Are there differences in the amount of talk from high- and low-ability students, from males and females, and so on?
3. What is the proportion of social/personal chat compared with "on-task" talk?
4. How many different students participate in discourse?

5. How much time is spent setting up and explaining the purpose of the lesson?
6. How much time is spent summing up and responding to student questions?

Example 3: Examining Practice Using a Self-Reflection Tool

The questionnaire presented in Figure 9.1 is intended to help teachers identify and reflect on their current and future professional development activities.

Example 4: Examining Practice or Knowledge Using a Self-Reflection Tool

Often, the simple task of rating oneself can lead to reflection about what one really knows or can do and what one needs to do to perform as well as one would like. In carrying out such self-ratings, it is better to rate specific areas of knowledge and skill than more general ones. The more specific the knowledge or skill, the easier it is to focus on its improvement, if necessary. Thus, rather than rating one's "assessment and evaluation skills" in general, an instrument such as the one shown in Figure 9.2 is useful for focusing on more detailed aspects of assessment and evaluation skills.

Example 5: Examining Teacher Effects Using Student Performance Data

This strategy is intended to promote teachers' thinking about both instruction and student learning. It is based on examination and reflection on students' learning gains. It is predicated on the availability of chapter or unit tests that can serve as a pretest and a posttest. Most textbooks provide both end-of-chapter assessments and review assessments. The latter can serve the function of a preunit test and the former the function of a postunit test.

The teacher should take the following steps:

1. Plan your instruction on the unit or chapter as you normally would.

Answer each question by circling the term that best describes your level of agreement, and then list current and planned activities in the spaces provided.

1. I am actively involved in developing my command of the knowledge base of teaching.

strongly agree *agree* *disagree* *strongly disagree*

Topics I an currently working on	*Topics I will work on next*
____________________	____________________
____________________	____________________
____________________	____________________

2. I am actively engaged in reflecting on issues related to myself as a teacher and my professional role.

strongly agree *agree* *disagree* *strongly disagree*

Topics I am currently reflecting on	*Topics I plan to reflect on next*
____________________	____________________
____________________	____________________
____________________	____________________

3. I am actively engaged in developing my repertoire of specific teaching and assessment practices.

strongly agree *agree* *disagree* *strongly disagree*

Practices I am currently learning	*Practices I plan to learn next*
____________________	____________________
____________________	____________________
____________________	____________________

4. What is the one area of teaching in which you feel you most need professional development? What evidence leads you to identify this particular need?

____

____

____

____

Figure 9.1. Sample of a Self-Reflection Tool

Rate your knowledge or skill as high, medium, or low for each of the following. For each skill, put an X under the appropriate rating category.

	Level of Knowledge or Skill		
Understanding or Skill	*High*	*Medium*	*Low*
My understanding of			
the effects of my grades on students			
biases in my grading procedures			
importance of grades to parents			
norm- versus criterion-referenced grades			
what my grades actually represent			
the fairness of my grading system			
My skill at			
diagnosing students' prior knowledge			
providing relevant feedback			
preparing students for my tests			
linking my teaching to my tests			
including nonrote items in my tests			
selecting valid textbook tests			
scoring essay items fairly			

Select one area of knowledge and skill in which you rated yourself low, and write a plan for how you could go about improving that knowledge or skill.

Figure 9.2. A Self-Reflection Tool for Detailed Assessment of Knowledge and Skills

2. Administer the pretest to students before instruction begins.
3. Score the tests not just in terms of total score, but by objective as well.
4. Based on the students' scores, determine (a) how well your original lesson plan fits the information from the pretest and (b) how you might have to revise the plan based on the pretest information; then, (c) make a prediction of how the class will do after instruction.
5. Teach the subject, making revisions in your plan as needed.
6. Administer the posttest and score it.

7. Compare the students' pretest and posttest performance, examining both total score and individual objectives.
8. On the basis of the comparison, ask yourself the following: (a) How did the class as a whole do relative to your prediction? (b) In what areas (objectives) did the class learn well, and which need improvement? (c) Why did the class not learn some areas as well as others? What might explain the discrepancy in learning across objectives? (d) What can I do next time to improve learning of those objectives?
9. Examine the performance of the three pupils who scored lowest on the posttest. What was the nature of their mistakes, and what could you do next time or in the next unit to help them improve?

Once the teacher has collected and examined the pre- and posttest information, many additional questions and issues will be raised.

Example 6: Examining Knowledge Using a Self-Reflection Tool

The self-evaluation instrument presented in Figure 9.3 is designed to help teachers rate their knowledge of some educational processes or activities.

What Are the Characteristics for Successful Self-Evaluation?

Although all forms of self-evaluation focus on teacher improvement, the most clear-cut aspect of fully implemented self-evaluation is its careful development and attention to detail at every stage of the self-evaluation process, from clear definition of the area to be evaluated to the collection and analysis of personal and external data. This form of teacher self-evaluation involves conscious planning and implementation, and generally leaves a paper trail of evidence (Kremer-Hayon, 1993; McColskey & Egelson, 1993). Other forms of teacher self-evaluation, such as reflection in or on action, generally do not (Hatton & Smith, 1995; Louden, 1992). However, although careful and conscious planning enhances the quality of self-evaluation, it also

For each process or activity, indicate how familiar you are with *how it could be applied in a classroom* by circling the appropriate response.

How familiar are you with the classroom implications and applications of

Alternative assessment	very familiar	somewhat familiar	heard the term	unfamiliar
Inclusion	very familiar	somewhat familiar	heard the term	unfamiliar
Multiple intelligences	very familiar	somewhat familiar	heard the term	unfamiliar
Portfolios	very familiar	somewhat familiar	heard the term	unfamiliar
Inquiry teaching	very familiar	somewhat familiar	heard the term	unfamiliar
Action research	very familiar	somewhat familiar	heard the term	unfamiliar
Advanced organizer	very familiar	somewhat familiar	heard the term	unfamiliar
Constructivism	very familiar	somewhat familiar	heard the term	unfamiliar
Metacognition	very familiar	somewhat familiar	heard the term	unfamiliar

Which two of these processes/activities do you most want to learn more about?

____

____

Where or to whom in your school or district would you go to get the information you want?

____

Figure 9.3. A Self-Reflection Tool for Examining Level of Knowledge

BEST COPY AVAILABLE

increases its complexity. There are, then, a number of characteristics that the successful self-evaluating teacher should possess and a number of obstacles that must be overcome.

Motive for self-evaluation. Successful self-evaluating teachers understand the need to take charge of their own practice and professional development (Hargreaves, 1995; Osterman & Kottkamp, 1993). Thus, as a starting point for self-evaluation, the teacher must have a commitment to practice and to his or her responsibility to understand, evaluate, and improve it. The initiative and desire for self-evaluation has to come from the teacher himself or herself. Principals, however, can directly encourage teachers to employ and model good self-evaluation practices.

Control of self-evaluation. If self-evaluation is truly to be *self*-evaluation, the teacher must assume control of the self-evaluation process. In particular, the teacher must be the one who carries out reflection and makes the ultimate decision about his or her practice.

Teacher control of self-evaluation does not preclude the possibility that others can help stimulate or initiate self-evaluation. A colleague's or administrator's viewing or reviewing aspects of the teacher's practice can be a powerful means of increasing the teacher's awareness of that practice (Anderson & Freiberg, 1995). Frequently, such awareness is created when an observer asks "why" questions. Why did you present the material that way? Why did you reject that student's argument when it seemed to have some merit? These externally generated questions focus on issues in the instructional process and may directly stimulate teacher reflection and self-evaluation.

Sharing control in the self-evaluation process can benefit the teacher, but it may also have undesirable effects. The benefits flow most directly from enhanced verifiability (e.g., when the teacher's perceptions are consistent with the perceptions of others) and the additional impetus given to self-evaluation. Undesirable effects result mainly from the teacher's perception of his or her own vulnerability in the evaluation situation.

Perceived ability to change practice. Unless a teacher acknowledges responsibility for or the possibility of being able to improve an area of practice, the self-evaluation process is unlikely to be carried through. Once a teacher becomes aware of a problem in practice, he or she must make a decision about the possibility of altering or improving it. Values, beliefs, prior experience, and available school support systems influence the teacher's perceptions and decisions about both the problem and the possibility of its melioration (Grimmett & Erickson, 1988). If the teacher believes that the factors that underlie a problem are outside his or her control (whether or not they really are) or not worth the effort to change, the self-evaluation process will cease. Many factors related to teachers' personal development and characteristics (e.g., risk-taking tolerance, self-efficacy, dogmatism) as well as the resources and support for professional development available in the school can influence teachers' interpretations of problematic classroom occurrences or situations (Ashton & Webb, 1986; Bandura, 1991; Nais, 1988). For this reason, the principal must help build a school climate that encourages and supports professional development through teacher self-evaluation.

Building a Support Structure for Self-Evaluation

Overcoming barriers to teacher self-evaluation is important and heavily dependent on both teacher and principal/administrator commitment. A school-based support structure should include the important features described below.

Awareness of self-evaluation and encouragement for teachers to conduct self-evaluation. The concept and processes of self-evaluation should be brought to the attention of teachers and encouraged as a schoolwide practice. Encouragement is important, because few teachers initiate and regularly use formal self-evaluation activities in their classrooms. The principal can help inform teachers about self-evaluation by making written resource information available to teachers, discussing self-evaluation in meetings, and setting expectations for the conduct of self-evaluation. The main objective of these activities is to introduce and sustain a positive environment for self-evaluation, one that supports self-evaluation and a general focus on the importance of teachers' responsibilities for professional development.

Assistance to teachers in learning about self-evaluation. Most teachers receive little instruction in assessment and self-evaluation practices. If these practices are to be implemented, some form of education for teachers is necessary. At a first level, teachers need to be provided with information about self-evaluation options, opportunities, and pitfalls. At a second level, teachers need to be informed about basic ideas and information for planning, organizing, and implementing self-evaluation at the school and classroom levels. At a third level, teachers can be given opportunities to experience the use and exploration of specific techniques and strategies for carrying out self-evaluation.

School policy that establishes self-evaluation as an integral part of the school structure. Few things are universally acclaimed and accepted from the outset. Most innovations need time and opportunity to become established and formal recognition to ensure longevity. Making self-evaluation an option within the formal structure of the school increases the relative importance of self-evaluation and enhances its long-term viability. Such recognition gives self-evaluation the exposure, standing, and credibility needed to encourage teachers to employ self-evaluation in their classrooms. Importantly, schoolwide guidelines can ensure that self-evaluation activities obtain the requisite resources, are conducted as part of the normal expectations for professional development in the school, are conducted appropriately, and are integrated with other aspects of instruction and evaluation.

Resources to help with and to encourage teacher self-evaluation processes. Self-evaluation can lead to its own reward of improved practice, but self-evaluation competes for time and attention against myriad other demands on teachers. How a principal chooses to allocate school resources sends an explicit message to teachers about what is valued and encouraged in the school. A clear indicator of the importance of any initiative is the willingness of the school administrator to apply resources to it. Resources can take the form of time, provision of physical facilities or aides, opportunities to visit with and learn from others, and a host of other possibilities.

Safeguards to protect teachers. Three safeguards will encourage teachers to participate in self-evaluation activities:

1. For self-evaluation to be most useful, it must be based on unbiased information. Bias is reduced when self-evaluation is conducted in an environment of support and trust, in which teachers can take an honest look at their practices without fear of sanction or penalty. It is also reduced when teachers learn to gather information that provides a basis for valid judgments about practice and when they obtain feedback from colleagues.
2. Self-evaluation is inherently risky. It is never easy to learn that there are weaknesses in one's practice or that one's beliefs about instruction are not borne out in the classroom. Negative information can be seized upon by others and used to the detriment of the self-evaluating teacher. School policy should provide guidelines to protect the rights and self-efficacy of teachers engaged in self-evaluation. Concrete commitments that ensure privacy of self-evaluation information and protect teachers who identify weaknesses in their practice must be provided at the school level.
3. Care needs to be taken to ensure the voluntary nature of self-evaluation. In particular, it is important to keep self-evaluation information and processes separate from externally imposed, summative evaluations. If voluntary and imposed teacher evaluations are mingled, teachers are likely to lose the feeling of voluntarism and the sense of personal empowerment that are fundamental to self-evaluation.

Conclusion

Teacher self-evaluation has the potential to help teachers examine and improve their practice in many ways. However, in applying the suggestions outlined above, it is important to recognize that each application will reflect the special cultures, values, resources, and personalities of the contexts and professionals involved. Part of the usefulness of teacher self-evaluation is its potential to spur examination of a school or teacher's context and practices to arrive at a strategy of teacher self-evaluation that best suits the setting of the involved teachers and administrators. If it is made meaningful to teachers and nurtured and supported by school administrators, teacher self-evaluation can give teachers a sense of ownership and pride in their practice and also improve instruction and learning in schools.

References

Airasian, P. W. (1993). Policy-driven assessment or assessment-driven policy? *Measurement and Evaluation in Counseling and Development, 26,* 22-30.

Airasian, P. W., & Gullickson, A. (1994a). Examination of teacher self-assessment. *Journal of Personnel Evaluation in Education, 8,* 195-203.

Airasian, P. W., & Gullickson, A. (1994b, Fall). Teacher self-assessment: Potential and barriers. *Kappa Delta Pi Record, 31*(1), 6-9.

Airasian, P. W., & Gullickson, A. (1995). *Teacher self-evaluation tool kit.* Report submitted to the Office of Research, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC.

Airasian, P. W., & Gullickson, A. (1996). *Teacher self-evaluation tool kit.* Thousand Oaks, CA: Corwin.

Allen, D. W., & Ryan, K. (1969). *Microteaching.* Reading, MA: Addison-Wesley.

Anderson, J. B., & Freiberg, H. J. (1995, Winter). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly,* pp. 77-91.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers.* San Francisco: Jossey-Bass.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Bailey, G. D. (1981). *Teacher self-assessment: A means for improving classroom instruction.* West Haven, CT: NEA Distribution Center. (ERIC Document Reproduction Service No. ED 207 967)

Bandura, A. (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Nebraska Symposium on Motivation: Vol. 38. Perspectives on motivation* (pp. 69-164). Lincoln: University of Nebraska Press.

Barber, L. W. (1987, January). *Teacher evaluation.* Report submitted to the Teacher Quality Program, Indiana Department of Education.

Barber, L. W. (1990). Self-assessment. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing* *elementary and secondary school teachers* (pp. 216-228). Newbury Park, CA: Sage.

Berliner, D. C. (1987). Ways of thinking about students and classrooms by more and less experienced teachers. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 60-83). London: Cassell Educational Limited.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Calderhead, J. (Ed.). (1988). *Teachers' professional learning*. Philadelphia: Falmer.

Carnegie Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the twenty-first century*. New York: Carnegie Forum on Education and the Economy.

Carter, K., & Doyle, W. (1987). Teachers' knowledge structures and comprehension processes. In J. Calderhead (Ed.), *Exploring teacher thinking* (pp. 147-160). London: Cassell Educational Limited.

Clandinin, D. J., & Connelly, F. M. (1988). Studying teachers' knowledge of classrooms: Collaborative research, ethics, and the negotiation of narrative. *Journal of Educational Thought, 22*(2A), 269-282.

Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher, 19*(2), 2-11.

Court, D. (1988). "Reflection in action": Some definitional problems. In P. P. Grimmett & G. L. Erickson (Eds.), *Reflection in teacher education* (pp. 143-147). New York: Teachers College Press.

Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 17-34). Newbury Park, CA: Sage.

Delandshire, G., & Petrosky, A. R. (1994). Capturing teachers' self-knowledge: Performance assessment. *Educational Researcher, 23*(5), 11-18.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5-12.

Duke, D. L., & Stiggins, R. (1985). *Teacher evaluation: Five keys to growth.* Washington, DC: NEA Professional Library.

Dwyer, C. A., & Stufflebeam, D. L. (1996). Teacher evaluation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 765-786). New York: Simon & Schuster/Macmillan.

Elbaz, F. (1990). Knowledge and discourse: The evolution of research on teacher thinking. In C. Day, M. Pope, & P. Denicolo (Eds.), *Insights into teachers' thinking and practice* (pp. 15-42). London: Falmer.

Eraut, M. (1993). Teacher accountability: Why is it central to teacher professional development? In L. Kremer-Hayon, H. C. Vonk, & R. Fessler (Eds.), *Teacher professional development: A multiple perspective approach* (pp. 23-43). Lisse, Netherlands: Swets & Zeitlinger.

Farnham-Diggory, S. (1994). Paradigms of knowledge and instruction. *Review of Educational Research, 64*, 463-477.

Feiman-Nemser, S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 505-526). New York: Macmillan.

Fuller, F. F., & Manning, B. A. (1973). Self-confrontation reviewed: A conceptualization for video playback in teacher education. *Review of Educational Research, 43*, 469-528.

Gadamer, H. (1975). *Truth and method.* New York: Seabury.

Gilliss, G. (1988). Schön's reflective practitioner: A model for teachers? In P. P. Grimmett & G. L. Erickson (Eds.), *Reflection in teacher education* (pp. 47-54). New York: Teachers College Press.

Gitlin, A., Bringhurst, K., Burns, M., Cooley, V., Myers, B., Price, K., Russell, R., & Tiess, P. (1992). *Teachers' voices for school change.* New York: Teachers College Press.

Glickman, C. D. (Ed.). (1992). *Supervision in transition: The 1992 ASCD yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Gore, J., & Zeichner, K. (1991). Action research and reflective teaching in preservice teacher education: A case study from the United States. *Teaching and Teacher Education, 7*, 119-136.

Grimmett, P. P., & Erickson, G. L. (Eds.). (1988). *Reflection in teacher education.* New York: Teachers College Press.

Grimmett, P. P., Rostad, O. P., & Ford, B. (1992). The transformation of supervision. In C. D. Glickman (Ed.), *Supervision in transition:*

The 1992 ASCD yearbook (pp. 185-202). Alexandria, VA: Association for Supervision and Curriculum Development.

Halkes, R., & Olson, J. K. (Eds.). (1984). *Teacher thinking: A new perspective on persisting problems in education.* Lisse, Netherlands: Swets & Zeitlinger.

Hargreaves, A. (1995). Renewal in the age of paradox. *Educational Leadership, 52*(7), 14-19.

Hargreaves, A., & Fullan, M. G. (Eds.). (1992). *Understanding teacher development.* New York: Teachers College Press.

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education, 11,* 33-49.

Haysom, J. (1985). *Inquiring into the teaching process: Towards self-evaluation and professional development.* Toronto: Ontario Institute for Studies in Education Press.

Iwanicki, E. F., & McEachern, L. (1983). Teacher self-improvement: A promising approach to professional development and school improvement. *Journal of Staff Development, 4,* 62-77.

Jones, A., & Airasian, P. W. (in press). Teacher self-assessment: Through the ages. *Midwestern Journal of Education Research.*

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62,* 129-169.

Kagan, N. (1975). *Interpersonal process recall: A method of influencing human interaction.* Mason, MI: Mason Media.

Kleinsasser, A. M. (1992). Learning how to teach language arts: A cultural model. In E. W. Ross, J. W. Cornett, & G. McCutcheon (Eds.), *Teacher personal theorizing: Connecting curriculum practice, theory, and research* (pp. 53-66). Albany: State University of New York Press.

Kottkamp, R. B., Provenzo, E. F., & Cohn, M. M. (1986). Stability and change in a profession: Two decades of teacher attitudes, 1964-1984. *Phi Delta Kappan, 67,* 559-567.

Kremer-Hayon, L. (1993). *Teacher self-evaluation: Teachers in their own mirrors.* Boston: Kluwer Academic Publishers.

Kuhn, D. (1991). *The skill of argument.* New York: Cambridge University Press.

Louden, W. (1991). *Understanding teaching: Continuity and change in teachers' knowledge.* New York: Teachers College Press.

Louden, W. (1992). Understanding reflection through collaborative research. In A. Hargreaves & M. G. Fullan (Eds.), *Understanding teacher development* (pp. 178-215). New York: Teachers College Press.

McColskey, W., & Egelson, P. (1993). *Designing teacher evaluation systems that support professional growth*. Greensboro, NC: Southeastern Regional Vision for Education.

McNamara, D. (1990). Research in teachers' thinking: Its contribution to educating student teachers to think critically. *Journal of Education for Teaching, 16*(2), 147-160.

Milner, J. O. (1991). Suppositional style and teacher evaluation. *Phi Delta Kappan, 72*, 464-467.

Nais, J. (1988). What it means to "feel like a teacher": The subjective reality of primary school teaching. In J. Ozga (Ed.), *Schoolwork: Approaches to the labour process of teaching*. Milton Keyes, UK: Open University Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Osterman, K. F., & Kottkamp, R. B. (1993). *Reflective practice for educators: Improving schooling through professional development*. Newbury Park, CA: Corwin.

Peterson, P. L., & Barger, S. A. (1985). Attribution theory and teacher expectancy. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 159-184). Hillsdale, NJ: Lawrence Erlbaum.

Pollard, A., & Tann, S. (1993). *Reflective teaching in the primary school* (2nd ed.). London: Cassell Educational Limited.

Ross, E. W., Cornett, J. W., & McCutcheon, G. (Eds.). (1992). *Teacher personal theorizing: Connecting curriculum practice, theory, and research*. Albany: State University of New York Press.

Saphier, J., & Gower, R. (1987). *The skillful teacher*. Carlisle, MA: Research for Better Teaching.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.

Schön, D. A. (1988). Coaching reflective teaching. In P. P. Grimmett & G. L. Erickson (Eds.), *Reflection in teacher education* (pp. 19-30). New York: Teachers College Press.

Schwab, J. (1969). The practical: A language for curriculum. *School Review, 78*, 12.

Sergiovanni, T. J., & Starratt, R. J. (1988). *Supervision: Human perspectives* (3rd ed.). New York: McGraw-Hill.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Silver, H. F., & Hanson, J. R. (1980). *Teacher self-assessment*. Moorestown, NJ: Hanson Silver & Associates.

Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *Politics of education association yearbook, 1990* (pp. 233-267). London: Falmer.

Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Boston: Kluwer-Nijhoff.

Tom, A. R. (1984). *Teaching as a moral craft*. New York: Longman.

Travers, R. (1983). *How research has changed American schools: A history from 1840 to the present*. Kalamazoo, MI: Mythos.

Valli, L. (1992). *Reflective teacher education: Cases and critiques*. Albany: State University of New York Press.

Van Manen, M. (1991). *The tact of teaching: The meaning of pedagogical thoughtfulness*. Albany: State University of New York Press.

Wertsch, J. V., & Toma, C. (1995). Discourse and learning in the classroom: A sociocultural approach. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 159-174). Mahwah, NJ: Lawrence Erlbaum.

Wildman, T. M., & Niles, J. A. (1987). Reflective teachers: Tensions between abstraction and realities. *Journal of Teacher Education, 38*(4), 25-31.

PART III

Implementing the Teacher Evaluation System

10

Conducting a Successful Evaluation Conference

VIRGINIA M. HELM

The evaluation conference in the school setting has for several decades tended to be tied directly to classroom observation, which occurred infrequently and, in its usual clinical supervision model, followed a highly prescribed pre- and post-observation format. The limitations of evaluating teachers solely or even primarily on the basis of once- or twice-yearly observations are, however, increasingly apparent. Questions of validity and reliability, of course, prevail, but so do concerns about subjectivity and even bias on the part of the supervisor. Furthermore, there is growing recognition of the need to evaluate the additional responsibilities and dimensions of teacher performance that go beyond a 1-hour classroom teaching exhibition.

Because so little has been written in the education literature about how to conduct an evaluation conference—except for the above-mentioned clinical supervisory pre- and post-observation conference—I have turned to the literature of business and the professions, where the evaluation conference (called a performance appraisal interview) is based on nearly three decades of theory, research, and practice. Since the 1980s and the advent of total quality management and Japanese-style participative management approaches that have shifted "control" from managers to employees —often in team settings—the traditional approach to evaluation has become anachronistic (Meyer, 1991). Peer involvement in work processes often leads to peer input in performance evaluation; emphasis on customer satisfaction sometimes leads to customer input in performance evaluation. Although the supervisor usually retains responsibility for developing the final, summary evaluation, he or she does so based on input from a wide variety of sources, often including self-appraisal by the employee being evaluated. Finally, the literature on performance appraisal in the professions and business world reflects a unanimous understanding of the evaluation conference as only one component in an ongoing communication and feedback cycle that is informal, timely, and specific (Buzzotta, 1988; Derven, 1990; Meyer, 1991; Nathan, Mohrman, & Milliman, 1991; Pennock, 1992). This feedback cycle shifts the supervisor's role from one of boss and judge (delivering an annual pronouncement about the employee's performance) to that of coach or mentor, where the supervisor is in frequent contact with the employee, offering regular —even daily—encouragement, praise, and constructive criticism immediately after observing specific actions or behaviors.

I should add that the education community has also begun to recognize the desirability of diversifying the range of materials and inputs submitted for the supervisor's consideration in constructing the final evaluation (see McGreal, 1988; Popham, 1988; Scriven, 1988). Finally, I want to note here that although the evaluation conference described in this chapter reflects the more contemporary approach to performance evaluation, much of what I discuss remains applicable to the standard clinical supervision conference as well.

What Are the Purposes of the Evaluation Conference?

The evaluation conference serves multiple purposes. As early as 1960, MacGregor (cited in Downs, 1990) identified three purposes that still hold true today:

- *Administrative:* to document performance for use in personnel decision making

- *Informative:* to inform the employee about his or her work performance
- *Motivational:* to motivate employees to higher levels of performance

The current literature suggests that a good evaluation conference can also serve *problem-solving, strategy-developing,* and *goal-setting* functions (Anderson & Barnett, 1987; Krayer, 1987; Losyk, 1990/1991; Meyer, 1991), as will be seen in the discussion that follows.

What has changed since 1960, however, when MacGregor identified the administrative, informative, and motivational purposes served by the evaluation conference, is that the conference is now understood as part of the ongoing feedback and communication that characterize good supervisory "coaching" (Somerick, 1993). Essentially, the evaluation conference confirms what has been communicated throughout the evaluation period. With regular feedback letting the teacher know where he or she has excelled and where there are concerns, there should be no surprises in the summary evaluation conference (Derven, 1990; Mancision, 1991; Pennock, 1992).

Preparing for the Evaluation Conference

The success of any evaluation conference depends on one factor that is simple and seemingly obvious, yet often seriously overlooked: careful preparation—not only by the principal but also by the teacher. Primary responsibility for planning, of course, rests with the principal.

Preparation begins with the principal setting a tentative time for the conference and checking with the teacher for any schedule conflicts. When selecting a location, the principal should look for a neutral conference room rather than using his or her office or the teacher's classroom. The principal should then provide the teacher with written notice of the date, time, and location of the evaluation conference, along with a copy of any evaluation form that is to be used and instructions to the teacher suggesting that he or she complete the form as part of a self-appraisal activity, which might also include the teacher's analysis of any external factors or conditions that have affected his or her ability to meet expectations. If the teacher has been keeping a portfolio or otherwise collecting data for evaluation purposes, he or she should also be asked to organize and prepare the material for submission—preferably several days to a week prior to the evaluation conference. The following list summarizes the eight steps a principal should take to prepare for an evaluation conference:

1. Set date, time, and place of evaluation conference after confirming with the teacher his or her availability at that time.
2. Give the teacher a copy of the evaluation form, requesting that he or she use it for self-appraisal.
3. Ask the teacher to organize, review, and submit any performance documentation collected.
4. Ask the teacher to be prepared to discuss successes, unmet challenges, what interferes with his or her best performance, and what the principal or school system can do to help the teacher achieve his or her goals.
5. Review any job description and previous evaluation along with any performance goals that were set for the evaluation period.
6. Complete a tentative evaluation and prepare notes summarizing the teacher's successes and concerns.
7. Plan a "script" for addressing concerns tactfully.
8. Prepare questions to enable the teacher to provide meaningful analysis of his or her strengths and areas for improvement.

If there is a job description for the teacher, the principal should review that carefully, along with a copy of the previous evaluation, to refresh his or her memory on any goals that were set during the last evaluation conference and any areas in which the teacher was expected to improve performance. If the principal has kept any documentation about the teacher's performance (critical incidents log, notes from informal drop-ins or observations, calendar), he or she should review those also.

Finally, the principal should tentatively complete the evaluation form used by the school or district. This will give the principal a clearer idea of the teacher's strengths and successes as well as any problem areas and unmet expectations or goals. Where there are deficiencies, it is crucial that the principal be prepared to articulate clearly yet tactfully the nature of the problems and some specific examples of behaviors or outcomes in which those deficiencies were exhibited. Additionally, the principal should consider some specific suggestions for improvement. The principal's jotting down all of these ideas in advance of the evaluation conference will ensure that he or she is focused and thorough during the conference, as well as enhance the likelihood of his or her communicating with tact when discussing any performance problems or deficiencies. If possible, the principal should conduct all of this review and appraisal several days or even weeks before the evaluation conference, to allow himself or herself time to step back and regain objectivity and distance on the situation.

The teacher also must invest some time and effort in preparing for the evaluation conference. Ideally, the principal will have provided the guidelines for such preparation, as described above, but in the absence of such guidelines, a teacher can maximize the benefits of the evaluation conference by taking the following steps:

- Collect, organize, and analyze any documentation generated during the evaluation period (sample assignments, tests, student work, pictures of displays, and so on).
- Review the job description (when available), previously set goals or objectives, and district or school mission statements.
- Complete a copy of the evaluation form provided by the principal.
- Identify major strengths and successes of the year.
- Identify any unmet expectations or goals and analyze possible reasons for failure to meet them, paying careful attention to factors both within and outside the teacher's control.
- Identify areas for growth (improvement or new directions) and possible goals or objectives for the next year.
- Identify how the principal or school system can help the teacher achieve greater effectiveness.

As the foregoing lists of activities necessary for effective preparation for an evaluation conference indicate, this conference is ideally an opportunity for reflection not only on the teacher's performance but also on the climate and conditions in which the teacher works. It can—and should—become an opportunity for joint problem solving and for the principal to obtain valuable insight from the teacher about the "health" of the school. And, as I emphasize throughout this chapter, the evaluation conference should not be an isolated, once-yearly event; rather, it should be a structured conversation and reflection that is part of the ongoing and usually less formal feedback process characterizing the principal's working relationships with teaching staff.

Conducting the Evaluation Conference

When the teacher arrives, the principal should do everything possible to establish a comfortable, relaxed atmosphere. To minimize authority barriers, the principal can sit next to the teacher, or both parties might sit in chairs on either side of an end table. The principal should have arranged in advance to remain uninterrupted by calls or visitors during the conference. Such prioritizing lets the teacher know how important he or she is and how important the conference itself is.

Much of the current literature suggests that the evaluator (in this case, the principal) invite the employee (the teacher) to share his or her self-appraisal first (Alexander Hamilton Institute, 1995; Buzzotta, 1988; Pennock, 1992). Research has demonstrated that evaluation conferences based on the employee's self-review "were significantly more constructive and satisfying to both parties than those based on the manager's appraisal. It also resulted in significant improvement in job performance" (Meyer, 1991, p. 72).

If the principal shares his or her evaluation first, and then asks, "What do you think?" the teacher is in an awkward position and may well respond with a shrug and the only viable reply: "I go along with that." This pretty well closes the discussion, precluding exploration of facts or issues. An additional advantage of allowing the teacher to speak first is that when the teacher has not met certain predetermined expectations or has experienced performance problems that he or she is willing to admit, the principal is spared the awkwardness of initiating such criticism. Because we are all more willing to accept self-criticism than criticism from others, allowing the teacher to identify problem areas first can also defuse some of the emotional content of the conference and reduce the natural defensiveness that interferes with receptivity to suggestions for improvement (Stronge & Helm, 1991). Clearly, if there are performance problems the teacher refuses to recognize, the principal must be willing to convey his or her concerns. If the principal has taken time in advance to prepare how to convey those concerns, this will greatly enhance the effectiveness of the evaluation conference.

To the extent that specific performance goals or objectives were set at the previous evaluation conference, the principal may open the conversation by asking the teacher to review the major goals and then indicate how close he or she came to accomplishing those goals. For example: "Eric, last year we agreed on two major goals you would pursue. Let's review what those were and the extent to which you believe you achieved those goals." Another opening might be to ask the teacher to summarize major strengths or major successes during the evaluation period: "What have you felt particularly pleased with in your classroom teaching this semester?" This question is open-ended but invites—even requires—specifics.

After focusing on and commending the teacher's strengths and accomplishments, the principal can comfortably shift the focus by asking the teacher in what areas he or she might have done a better job or might be able to do a better job in the coming year. At this point, the principal may also ask what he or she can do to help the teacher achieve success in those areas cited for improvement.

In situations where the teacher has not attained a specific goal or level of performance, the principal should ask for an explanation. For example:

> *Principal:* Cheryl, one of your goals for the year was to send 50% fewer students to my office for disciplinary action. How close do you think you came to reaching that goal?
>
> *Teacher:* Well, Ms. Principia, I wasn't counting, but I'm afraid I didn't make the kind of improvement I had hoped to make.
>
> *Principal:* What factors do you think made progress slower than you expected?
>
> *Teacher:* I've given it some thought in preparation for this evaluation conference, and I think part of the problem was that I had a much larger number of at-risk students in my class this year. Instead of 5 out of 25, I had 12—that really makes a difference, you know.
>
> *Principal:* Yes, Cheryl, I understand that it does. You didn't come to me for any additional assistance—did you contact the special education consultant or teacher for some suggestions for behavior modification?

Here, the principal shows that she has listened and understands, but also conveys awareness that the teacher bears some responsibility for taking initiative to solve her own problems.

Admittedly, explanations may sometimes consist of excuses, in which case the principal must prod the teacher to go beyond making excuses, and even be prepared to offer alternative explanations, preferably in question form. For instance: "Ed, your not hearing your alarm clock isn't sufficient to justify your frequent tardiness. What other alternative ways can you think of to get yourself to school on time?"

Sometimes, however, there will be times when legitimate mitigating circumstances, such as those in Cheryl's scenario above, should be taken into account in the final evaluation. Furthermore, the principal's asking the teacher to discuss the context and external or environmental factors affecting his or her performance may reveal some important information about the school or classroom environment that can be the basis for problem solving to remedy the situation.

What Are the Characteristics of Effective Evaluation Conferences?

The literature of performance appraisal reveals several components of effective evaluation conferences. Specifically, the five elements described below are central.

Two-way communication. Numerous studies have shown that the more an employee (i.e., the teacher) participates in the evaluation conference, the more satisfied he or she will be with the conference and the supervisor (principal) (see, e.g., Anderson & Barnett, 1987; Krein, 1990; Webb, 1989). Principals who listen more than they talk will obtain more useful information about teachers' performance and development needs. A principal's encouraging substantial contributions by the teacher also ensures greater commitment of the teacher to the process, which in turn improves the likelihood of success during the next evaluation cycle.

Balanced review of past performance and future goals. We in education have too frequently focused on past (or, with clinical supervision, present) performance. As we move toward more emphasis on setting performance goals or developing professional growth plans, we must necessarily begin devoting more of the evaluation conference to development of those goals and plans for future professional growth. Again, applying lessons taken from the literature, principals are well-advised to encourage teachers to take the initiative in identifying the goals and plans that are most meaningful to the school. This serves a dual purpose: (a) It ensures ownership of the goals selected for the teacher to pursue during the next evaluation period, and (b) it contributes to the teacher's positive perceptions about the evaluation conference itself (Downs, 1990).

Recognition of teacher strengths and successes. This component is one that hardly needs repeating, given that principals already realize the impact of such recognition on motivation and morale. What is important to remember, however, is that the more specific the commendations, the more meaningful they are to the receiver. Keeping a "critical incidents log" of particular events or successes throughout the year can aid the principal in developing a list of examples to reference when making generalized comments. (Presumably, the principal also will have commended the teacher at the times the successes occurred.)

Identification and analysis of problems affecting the teacher's performance. Few teachers will meet or exceed all goals or expectations set the previous year. Where there are disappointments (failures) in meeting those expectations, the principal should encourage the teacher to first identify and then analyze the reasons for these unmet goals or performance expectations. The purpose is to understand the context for these failures and then to move immediately toward identifying solutions to the problems that interfered with maximum effectiveness. The principal should be especially quick to ask what he or she can do to help resolve the problem(s) identified. The principal can also ask what the teacher believes he or she can do to contribute to solving the problem. Realistically, I should add that where the principal has identified performance problems overlooked or unconfessed by the teacher, he or she will, of course, need to present those concerns to the teacher tactfully and then pursue the joint problem solving discussed earlier.

Teacher initiation of goals for the next evaluation cycle. Allowing the teacher to be actively involved in the goal-setting process has been shown to be a more powerful catalyst for later performance than a general discussion about goals (Anderson & Barnett, 1987). Creating a sense of ownership in the goals or job expectations may be the most potent factor, not only in commitment to those goals but also in the teacher's perception of the success of the evaluation conference (Klein & Snell, 1994). The principal can facilitate the teacher's goal setting by asking questions that derive from the conversation about achievements and areas for growth (which might be areas for improvement or development of new skills/knowledge) earlier in the conference. Again, the principal should be prepared to offer a couple of goals on those occasions when a teacher is unwilling or unable to suggest some.

To reiterate, the characteristics of an effective evaluation conference can be summarized as follows:

1. Two-way dialogue, a majority of which is teacher talk
2. Balanced review of past performance and planning for the next evaluation cycle
3. Principal's recognition of the teacher's strengths and successes
4. Identification and analysis of problems affecting the teacher's performance
5. Teacher initiation of goals for performance improvement for the next evaluation cycle

How Can Principals Facilitate Productive Teacher Talk?

As I mentioned earlier, the more the teacher contributes to the evaluation conference, the more likely the event will be perceived by both principal and teacher as effective. A good talking ratio has been suggested as 80% teacher, 20% principal (Alexander Hamilton Institute, 1995).

I have already suggested as good openers (after any initial small talk) questions that elicit the teacher's perception of his or her strengths and successes or the teacher's review of previously established goals/objectives and the extent to which he or she has achieved them. Some teachers, however, may be reticent about speaking in the conference, whether because of shyness, anxiety, or possibly even resentment. A principal who responds to such reticence by doing all of the talking, however, will be left worse off than before the conference, and so will the teacher. No productive dialogue having occurred, the teacher will feel no ownership in the event or its outcome, nor will that teacher have any reason to be committed to any goals or expectations set solely by the principal. How can a principal get such a teacher to talk? A number of useful strategies are discussed below (Alexander Hamilton Institute, 1995).

Silence. Silence can be effective because it is not judgmental or offensive. When the teacher makes a negative statement (e.g., "This class is making me crazy" or "I think these evaluations are a waste of time"), a response of silence by the principal invites the teacher to say more. If the principal can wait in silence, the teacher will often go on to provide details. By not responding judgmentally, the principal does not put the teacher on the defensive. It takes a lot of fortitude, however, to use this strategy, because we are all conditioned to fill silences. If the principal can let that conditioning compel the teacher, however, he or she may gain valuable information as well as be perceived as an excellent listener.

Staying on track. Although it is true that I have emphasized the importance of teacher talk, it is equally important that the teacher talk about performance-related issues. Thus, the principal is responsible for directing the conversation and keeping it going. He or she must continually probe for specific information and ideas, in part by avoiding questions that produce "yes," "no," or "I guess so" answers. The ques- tion "Do you think you're doing a good job in the classroom?" is a start, but it is not as effective as "How do you feel you're doing in the classroom?" Even more effective would be such questions as "What aspects of teaching do you feel you do especially well?" and "What do you do in the classroom that you feel especially good about?" These questions invite not only a list of several behaviors but almost inevitably a brief discussion of the importance or results of those behaviors. Closed-ended questions producing all but useless answers tend to begin with "Did you," "Will you," or "Can you," as well as "How many" or "When."

Open-ended questions. Open-ended questions invite expansive, thoughtful, and informative answers. For example, if the principal asks, "Why do you think you're having trouble keeping students in their seats?" the question encourages the teacher to state the cause, making him or her less defensive than if the principal had offered the same explanation. The principal can then go on to ask, "How would you suggest we deal with this?" This question makes teacher and principal partners in solving the problem.

Drawing out explanations. When a teacher gives a cursory answer, the principal can invite further detail by making such requests as "Give me the details," "Please explain that," "Say a little more," or "Tell me a little more about that." To invite somewhat more thoughtful reflection on a teacher's decisions, a principal might ask, "How did you happen to choose that strategy?" or "What was your rationale for choosing that approach?" Usually the most reticent teacher will be able to continue with such prodding, but the principal may still need to exercise the silence strategy to accomplish the goal of obtaining a fuller explanation.

Presenting a problem. Another way a principal might obtain an open-ended response is to present a real or hypothetical problem to the teacher and ask for a possible solution. "Suppose I were to say that we need to reduce the number of students you send to the office by 50% next year. What might each of us do to accomplish this goal?" An alternative might simply be to establish the goal as a schoolwide goal or as a goal for a hypothetical teacher. Such questions not only help the principal assess the teacher's approach to problems but also lead to realistic goal setting for the next year.

Active listening. Active listening goes beyond the silence approach discussed above, and it takes a variety of possible forms. Responsive listening entails both verbal and nonverbal responses, such as nodding the head, smiling, or saying "I see," or "I understand." Empathic listening, on the other hand, entails restatement to the speaker, indicating that one understands what he or she has said. For example, the teacher may say, "The worst part of the job is grading exams and projects. The reason my semester grades are late so often is because I hate doing it so much." The principal might then reply, "You really find it that unpleasant to do?" No judgment has been communicated, but the teacher knows the principal has heard and understands the teacher's feelings about the task. Conveying judgment, advice, or sarcasm may be less effective. For instance:

- *Judgment:* "We need those grades turned in on time so that all the report cards can be sent out on schedule."
- *Advice:* "You'd better start getting those grades in on time or you're going to be in serious trouble."
- *Sarcasm:* "I thought you must enjoy grading, since you are so reluctant to end the process by turning the grades in."

These kinds of responses are guaranteed to make the teacher defensive and resistant to suggestions or recommendations for improvement.

Other listening skills. Following are a few additional suggestions for principals to improve their genuine listening during the evaluation conference (Alexander Hamilton Institute, 1995):

- Remember to talk less. You'll learn more.
- Don't make hasty judgments. Seek instead to probe for real reasons behind problems or behind stated explanations.
- Read the nonverbal cues. When there is a discrepancy between what the teacher says verbally and what the body language indicates, the principal should inquire directly about that discrepancy.
- Don't interrupt. Even if the principal has figured out what the teacher is going to say, it is rude and conveys disinterest in the teacher to finish the sentence for him or her or to interrupt with a response before the teacher has finished.
- Acknowledge ideas. There is a difference between accepting and acknowledging a teacher's idea. Acknowledging the idea, and even going beyond to state its benefits or impacts, conveys the principal's receptivity to ideas, which will help ensure continued sharing of ideas.
- Ask for clarification. An excellent way of communicating to a teacher that his or her comments are being carefully listened to is to ask for clarification or expansion where the description

is vague or general. Such follow-up can produce valuable recommendations or suggestions.

Use of pronouns. It is important for evaluators to be aware of possible evaluatee responses to statements dealing with performance problems. Generally, the literature suggests avoiding *you* and *your* statements, which can sound judgmental and make people defensive. For example, the statement "Your grading standards are too unrealistic" will only make a teacher resistant to suggestions for change. An alternative approach might be to say, "Many of your students feel that your grading standards are so high that there is no point in trying." Here, the principal places the emphasis on the students' perceptions. This might lead into an open-ended question or request for an explanation designed to get at the root of the problem: "I appreciate any teacher who sets and holds to high standards for their students. It is not the easy thing to do. But I would be interested to learn your rationale for setting standards so high that no student earns an A." Now the teacher has an opportunity to reflect on why he or she gives lower grades than other teachers, and the principal may be able to probe further by asking how effectively this approach enables the teacher to achieve the expressed goal.

Some authorities suggest replacing *you* statements with *I* statements, but it seems there is a spectrum from *you* through *I* to *we*, or to a *we* question. "You ought to start your classes as soon as the bell rings" places the emphasis on the teacher and what he or she should do. "I want you to start your classes as soon as the bell rings" changes the emphasis to the principal and what he or she wants. But an even more effective approach would be to say, "We all benefit when teachers start their classes on time. Is there anything we on the staff can do to help you do this more often?"

Focus on results or impact. In the case of a teacher who frequently starts classes late, the principal might explain, "When you start your classes as soon as the bell rings, your students will be more likely to arrive on time, get in their seats, and spend more time on task. It's a good way to communicate your expectations and minimize student disruption." Or in the case of a teacher who consistently turns in grades late, the principal might ask, "Have you ever thought about the consequences to the system when a teacher turns grades in late?"

By shifting the focus from the teacher's behavior to the impact or results of that behavior, the principal enables the teacher to look at the issue more objectively and less defensively.

Concluding the Interview

Principals hardly need reminding of the importance of concluding an evaluation conference on a constructive, encouraging note. After summarizing the teacher's main accomplishments and strengths and reviewing one or two areas targeted for professional growth in the next year, the principal should conclude with a couple of additional statements or questions designed to bring closure to the interview and to motivate the teacher toward continued dedication. For example, the principal might ask, "Carol, before we finish, do you have any questions about the appraisal or about our expectations for you next year?" Then, the principal might conclude,

> Carol, you make an invaluable contribution to our school community. Your dedication and your competence are recognized by many others besides myself. You're to be commended for your willingness to try out the new science curriculum next year. If you run into any difficulties, I hope you'll drop by—you know I'll provide whatever support I can.

Or, if it has been necessary to emphasize some areas for remediation, the principal may reshape the conclusion along these lines:

> Kent, your subject-area competence is unquestioned. Your students like you, and I know you have unrealized potential at this point. As you work on classroom discipline, please feel free to contact me for additional help, or chat with one of your colleagues down the hall. I am confident that I'll be seeing far fewer of your students in my office next year.

Summary

The most successful evaluation conference is a dialogue between principal and teacher, not a monologue by the principal. That dialogue, in fact, should contain much more teacher talk than principal talk. This conference is a more structured component of an ongoing supervisory system that approximates the coaching function, with regular, immediate, and specific feedback. Both principal and teacher must spend time reviewing and reflecting on the previous evaluation period in order to maximize the time spent in conferencing. Discussion is not limited to the "what" of teacher performance but also explores the "why" and "how" of that performance; it also focuses on the larger context in which the teacher's performance must be appraised. As such, the conference can also become an excellent forum for problem solving, as teacher and principal identify constraints on or obstacles to the teacher's maximum performance. The focus is not limited to the past, however. It should also include significant emphasis on the future and on any areas the teacher or principal has identified as new opportunities for professional growth or improvement. Finally, the evaluation conference should make clear that the principal is committed to providing assistance and support as the teacher undertakes new challenges or attempts to correct performance deficiencies.

An evaluation conference characterized by the features described above will be successful because the teacher will have participated as a professional, will feel appreciated and valued, and will have offered many of the major insights and much of the direction for the coming year. The principal and teacher will leave the conference with renewed appreciation for each other and fresh awareness of a mutual commitment to the school community and the learning enterprise.

References

Alexander Hamilton Institute. (1995). *Effective interviews for every situation.* Ramsey, NJ: Modern Business Reports.

Anderson, G. C., & Barnett, J. G. (1987). Characteristics of effective appraisal interviews. *Personnel Review, 16*(4), 18-25.

Buzzotta, V. R. (1988). Improve your performance appraisals. *Management Review, 77*, 40-43.

Derven, M. (1990). Assessment: The paradox of performance appraisals. *Personnel Journal, 69*, 107-111.

Downs, T. M. (1990). Predictors of communication satisfaction during performance appraisal interviews. *Management Communication Quarterly, 3,* 334-354.

Klein, H. J., & Snell, S. A. (1994). The impact of interview process and context of performance appraisal interview effectiveness. *Journal of Managerial Issues, 6,* 160-175.

Krayer, K. J. (1987). Simulation methods for teaching the performance appraisal interview. *Communication Education, 36,* 276-283.

Krein, T. J. (1990). Performance reviews that rate an "A." *Personnel, 67*(5), 38-41.

Losyk, B. (1990/1991). Face to face: How to conduct an employee appraisal interview. *Credit Union Executive, 30*(4), 24-26.

Mancision, J. (1991). The appraisal interview: Constructive dialogue in action. *Health Care Supervisor, 10,* 41-50.

McGreal, T. L. (1988). Evaluation for enhancing instruction: Linking teacher evaluation and staff development. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 1-29). Alexandria, VA: Association for Supervision and Curriculum Development.

Meyer, H. H. (1991). A solution to the performance appraisal feedback enigma. *Academy of Management Executive, 5,* 68-76.

Nathan, B. R., Mohrman, A. M., Jr., & Milliman, J. (1991). Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: A longitudinal study. *Academy of Management Journal, 24,* 352-369.

Pennock, D. (1992). Effective performance appraisals (really!). *Supervision, 53*(8), 14-16.

Popham, W. J. (1988). Judgment-based teacher evaluation. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 56-77). Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1988). Evaluating teachers as professionals: The duties-based approach. In S. J. Stanley & W. J. Popham (Eds.), *Teacher evaluation: Six prescriptions for success* (pp. 110-142). Alexandria, VA: Association for Supervision and Curriculum Development.

Somerick, N. M. (1993). Strategies for improving employee relations by using performance appraisals more effectively. *Public Relations Quarterly, 38*(3), 37-39.

Stronge, J. H., & Helm, V. M. (1991). *Evaluating professional support personnel in education*. Newbury Park, CA: Sage.

Webb, G. (1989). Room for improvement: Performance evaluations. *Wilson Library Bulletin, 63*(6), 56-57, 125.

Additional Sources

Armentrout, B. W. (1993). Eight keys to effective performance appraisals. *HR Focus, 70*(4), 13.

Austin, N. K. (1992, November). Updating the performance review. *Working Woman, 17*, 32-35.

Baker, J., Jr. (1988). *Causes of failure in performance appraisal and supervision: A guide to analysis and evaluation for human resources professionals*. Westport, CT: Quorum.

Bushardt, S. C., Jenkins, J. M., & Byrd, P. C. (1990). Less odious performance appraisals. *Training and Development Journal, 44*(3), 29-35.

Denton, D. K. (1987). How to conduct effective appraisal interviews. *Administrative Management, 48*(2), 15-17.

Gates, A. (1991). The smartest way to give a performance review. *Working Woman, 16*(5), 65, 68.

Goodale, J. G. (1992). Improving performance appraisal. *Business Quarterly, 57*(2), 65-70.

Ilgen, D. R., Peterson, R. B., Martin, B. A., & Boeschen, D. A. (1981). Supervisor and subordinate reactions to performance appraisal sessions. *Organizational Behavior and Human Performance, 28*, 311-330.

Kirkpatrick, D. L. (1982). *How to improve performance through appraisal and coaching*. New York: AMACOM.

Latting, J. K. (1992). Giving corrective feedback: A decisional analysis. *Social Work, 37*, 424-430.

MacLean, J. C. (1994). Personnel evaluation in education: A model for job-specific criteria. *Clearing-House, 68*(1), 47-51.

Marien, B. A. (1992). Quality in R&D: Putting TQM into the performance review process. *Research-Technology Management, 35*(5), 39-43.

McBey, K. (1994). Perfecting performance appraisals. *Security Management, 38*(11), 23-25.

Newman, R. J. (1994, November 1). Job reviews go full circle. *U.S. News & World Report*, pp. 92-98.

Patten, T. H., Jr. (1982). *A manager's guide to performance appraisal*. New York: Free Press.

Plachy, R. J. (1988). *Performance management: Getting results from your performance planning and appraisal system*. New York: AMACOM.

Plachy, R. J., & Plachy, S. J. (1993). Focus on results, not behavior. *Personnel Journal, 72*(3), 28-33.

Rausch, E. (1985). *Win-win performance management/appraisal*. New York: John Wiley.

Sandler, L. (1990). Two-sided performance reviews. *Personnel Journal, 69*, 75-78.

Umiker, W. (1992). Performance review interviews: Planning for the future. *Health Care Supervisor, 10*(3), 28-35.

11

Dealing Positively With the Nonproductive Teacher

MARY JO McGRATH

Excellence in education is our nation's call to arms. Yet accountability for producing that result is sending shock waves through the field. The demand for accountability is an explosive dynamic involving both organizational systems and individual behavior. Called into play are issues of personal responsibility for our own actions, the discomfort of being the one to hold others accountable for performance results, and the paramount responsibility of answering to parents and the community for the quality of our children's education.

The American judicial system has a long history of wrestling with accountability, as it determines what is appropriate and fair when passing judgment on a teacher's continued employment. From extensive inquiry, the courts have delineated practical criteria that provide sound guidance on what behaviors, on the part of both employees and management, determine quality performance and powerful administrative leadership.

Equally as important as these practical criteria governing actions is an assessment of the ethics of the culture and climate in which those actions are taken. For most people, *accountability* is a frightening word, implying "feet to the fire for failures." The fear people experi-

experience around accountability elicits blaming behavior to avoid responsibility, even among the most committed and ethical people.

For true progress to occur and excellence to be reality, these two tracks, practical criteria and workplace ethics, must merge into a cohesive management approach that balances the whole. Competent administration of successful schools must include both the skill and the courage to manage for the utmost success of students and staff.

In this chapter, I present the criteria for accountability and their companion ethical factors, and discuss how they should be incorporated into the day-to-day practices of school administrators. If this holistic approach is adopted, educators are much more likely to succeed in their efforts to be accountable and to realize the goal of quality education for all students.

How Do We Make Accountability an Opportunity, Not a Punishment?

The decision to terminate a permanent teacher's employment based on faulty performance is reached after a multitude of events have taken place. These events involve all aspects of the school community, including administration, students, peers, and parents. Many aspects of the legal process are involved, often including lawyers, a trier of fact, and judicial review.

An assessment of the appropriateness of a teacher's dismissal also encompasses an analysis of many personal influences in the teacher's life: his or her health, marital difficulties, extended family complications, and the like. Particularly nerve-racking for the principal, teacher dismissal includes a close look at the competence of the administrator in the areas of leadership, supervision, documentation, and evaluation. The legal criteria that have arisen around this area of employment law have evolved to weigh and balance the whole, so as to enable reaching a judgment on appropriate action.

Teacher dismissal opens the door to scrutiny of the entire person of the teacher and the entire school site's functioning. Competent performance of an administrator in the area of evaluation and accountability is not only a matter of observing classroom performance, accumulating input from portfolios or other means, and writing comprehensive summary evaluations. Administrator competence is a matter of being attuned to the entire organism that constitutes a school, rather than performing disjointed assessment functions. As Marshall (1995) notes, "The unexamined application of Newtonian laws to social systems caused us to design . . . linear system(s) . . . when in fact human systems are complex, dynamic and organic" (p. 11). The multidimensional, dynamic nature of educational institutions is recognized in the law, and the legal criteria are designed to consider the interrelationships within the whole.

This systemic approach can be viewed as a predicament or an opportunity. When professional administrators realize that the judicial process surrounding teacher dismissal considers the whole of the situation from an educational, legal, and ethical perspective, they may be stimulated to learn what factors are taken into account and to make attention to those factors part of their everyday repertoires. The highest measure of administrator competence is the fair, ethical, and professional treatment of employees incorporated into the day-to-day practices of administrators in holding people accountable for excellence in education.

How Does the Law of Accountability Consider the Whole?

Typically, we hear the phrase *the law* and we think two things: Trouble is brewing, and we are going to get tied up doing things that are irrelevant to getting the real job of educating students done. Recently, one educator challenged the usefulness of the law as applied to education, pointing out that the legal system seems to be the only societal system in more trouble than education. In reality, the law of accountability applied to teacher evaluation and termination is very broad, flexible, and—believe it or not—user-friendly.

It is important to distinguish between the procedural aspects of the law, which apply to such things as hearing rights and appellate review, and the criteria that have evolved within the law related to the actual components of a teacher's performance and employment. Whereas the procedural protections can seem to be a ghastly maze, the substantive guidelines are based in thoughtful reflections that have emerged over time as courts have confronted poorly performing teachers and their impacts on the students they educate. The judicial standards are in tune with the fundamental essence of the law, which is meant to address the question "When people decide they are going to coexist, what should the rules be so that everyone thrives?" Viewed in this light, the legal criteria may be seen as a fit for educator temperament, which most often is that of nurturing caregiver, not quasi-lawyer and case builder.

There are three legal criteria that govern the area of teacher performance and retention: due process, just cause, and fitness for service. These concepts are used collectively to determine the appropriateness of the supervision and evaluation processes and, in the cases where necessary, whether a teacher should be removed from the classroom and his or her employment terminated. What is even more important is that these criteria provide excellent guidance on the smooth functioning of the school as a dynamic organism. These legal concepts are discussed below in terms of their technical definitions and their translation for everyday use by administrators.

Due Process = Don't Blindside People

The assessment of the presence of due process on a day-to-day basis involves asking the question "Is the teacher receiving notice of problems and an opportunity to correct them?" In legal terms, due process requires that a person receive notice of the charges against him or her, an opportunity to review any documentary evidence, and the chance to respond (Alexander & Alexander, 1985; *Cleveland Board of Education v. Loudermill*, 1983). The practical essence of this legal concept is to give employees a contemporaneous opportunity to be informed of and to correct deficiencies when they are originally noted by the administrator. Said in common parlance, people should not be surprised and blindsided by deficiencies that show up in their summary evaluation reports.

An administrator should avoid at all costs the practice of creating personal files on employees; files filled with anecdotal notes, records of things the administrator has observed or heard about concerning employees as they go about their duties. Often an administrator will fail to inform an employee that such information exists until a situation deteriorates. Sometimes an employee will see a reference to a complaint based on behaviors observed many months earlier, or a past parent complaint, for the first time during a final evaluation. If an administrator is keeping such a file, the teacher has the right, legally and ethically, to know about it and its contents.

In order to observe fairness and not surprise an employee after the fact, administrators must follow these four steps in the maintenance of any tangible material on the employee:

1. The employee must be informed that the material exists and is being retained.
2. The employee must be informed of his or her right to comment in writing and have the comment included with the original material.
3. This notification must be done in a timely fashion, meaning within a reasonable amount of time so that the employee can remember pertinent facts about the event recorded and respond to the concerns stated.
4. The source of the information must not be anonymous or unidentified to the employee (*Miller v. Chico Unified School District*, 1979). (It is acceptable for client surveys [e.g., parent/student surveys] to be designed to be anonymous.)

Just Cause = Be Fair

Just cause fundamentally means "Be fair." This legal concept has been interpreted by arbitrators and judges as requiring in all cases an assessment of the totality of the circumstances. In any individual case of contested evaluation or termination, everything about the situation is taken into account, from employee conduct to management response. Key questions asked in judicial review are as follows:

1. What did the teacher do or not do?
2. How did those behaviors affect the school community, including pupils, coworkers, staff, administrators, and parents?
3. What context was the teacher working within? (a) Were the supervisor and school district acting in good faith, clearly pointing out problem areas and diligently and honestly striving to assist in improvement? (b) Were there unforeseen difficulties operating as extenuating circumstances that need to be considered?

The courts have ruled that for determining the propriety of the termination of an employee, the terms *just cause* and *good cause* have been difficult to define and depend on the circumstances of each case. Essentially, these terms connote a fair and honest cause or reason, regulated by good faith on the part of the party exercising the power (*Moore v. May Dept. Stores Co.*, 1990). The actions of the employee alone are rarely enough to warrant termination. They are considered in the context of the actions of management.

Fitness for Service = Don't Be Nitpicky

The fitness for service criterion takes into account many different factors that, when taken together, answer the question "Was the disruption significant to the educational process or was it inconsequential?" (See *Morrison v. State Board of Education*, 1969.) If the answer is that it was inconsequential, the likelihood that an evaluation can be thrown out is high, and the likelihood that dismissal will occur is low.

In determining fitness for service, courts take a look at matters that actually show that a teacher should not continue in the classroom because of the harm being generated (*Bassett Unified School District v. Commission on Professional Competence*, 1988; *Morrison v. State Board of Education*, 1969). These factors fall into the eight categories described below.

Harm caused by the conduct. The courts look to see whether the conduct engaged in by the teacher actually resulted in harm, that is, disruption of the educational process or impairment of teacher-student relationships. Harm may be detected in several areas—not just to students, but also in negative impact on the school district, the teachers, classified staff, administration, and/or the community. Without a showing of harm, it is nearly impossible to terminate a teacher.

On a regular basis, it is important for the administrator not only to specify the behavior a given employee is engaged in that is considered less than satisfactory, but also to track that behavior to its impact on students, staff, and others. The administrator should ask himself or herself, "What do I say, in my professional opinion, is the impact on students, teachers, staff, administration, community, parents, or others?" He or she should then record the answers in the documentation, along with a specific description of the behaviors observed. This factor is governed by the principle that having an employee be considerate of his or her impact on others is essential to the employee's understanding that what he or she does affects other people and matters to the smooth functioning of the whole.

Notoriety of conduct. Courts may look at the notoriety of the teacher's conduct in considering the seriousness of any disruption caused by the teacher's behavior. What this means on a practical level is that the administrator should determine whether the improper conduct is known to others and whether it is disrupting the teacher-student or teacher-parent relationship or impairing the reputation of the school district. Here the question the administrator should ask is "What are the comments from others (coworkers, faculty, administration, community, students, parents, public) about the impact or effects of this behavior/incident/pattern of conduct?" The guiding ethical principle is that people are accountable for the reputations they generate by their actions and are answerable to others for their effectiveness. The administrator must be sure that complaints received are not secreted away, but shared contemporaneously with the employee. Also, the administrator must follow any parent complaint procedure established in his or her school district.

Proximity or remoteness in time. The courts look to see whether the conduct charged against an employee is recent. If the conduct occurred so long ago that it can be said to be stale, the remoteness of the conduct will argue against its inclusion in the evaluation or dismissal of the employee. Note that if the administrator responds to situations in a timely fashion, the issue of proximity in time will not be a problem. Being timely with the communication of problems to the employee goes along with the notice requirements detailed above in regard to the retention of documentation. To check themselves on timeliness, administrators should ask the question "What is today's date? Is it near in time to the incident/conduct at issue?" The ethical principle here involves the administrator's effort to be aboveboard in the management of issues, disciplining himself or herself to handle things as they come up, and not procrastinating.

Extenuating or aggravating circumstances. The fitness for service criterion examines whether there is something happening surrounding the activity of the teacher that would either explain the conduct or be an additional aggravating circumstance, rendering the conduct even more of a problem. An example of the former would be if an employee's spouse has just suffered a very severe illness or the loss of a job. An example of the latter would be a case in which the employee has been continually advised of the requirement to be at school on time and there has been continuing failure to improve performance. The question the administrator should ask here is "Were there any extenuating circumstances (identified by the administrator or the employee) that are external to the employee that are affecting performance?" The principles of awareness and compassion should lead managers to make judgments that take into account the whole of the circumstances and factor them into the solution.

Motive for conduct. It is crucial to determine early in the process whether the teacher has any attitudinal justification for how he or she is behaving. If this factor is ignored, at the time of trial an understandable "excuse" may be given that undermines the district's ability to obtain a dismissal. The employee should be asked at the investigative conference regarding the conduct why he or she has engaged in the behavior, and any response should be recorded in the conference memorandum. The administrator should answer the question "What statements or conduct reflects the internal motivation or attitude of the individual about this conduct/incident/pattern of conduct?" On an ethical basis, the administrator is assessing the employee's will to address the issue and whether he or she can be counted on to bring forth a commitment to its resolution.

Prior help given. An important factor is whether an employee has been notified of the improper conduct *and* provided assistance in an effort to remediate the problem. A key aspect here is whether the teacher made any effort to correct the situation, particularly after receiving help, advice, or direction from the administration. Also important is whether the administration has provided assistance appropriate to correction of the deficiencies; usually, this is understood to include the district's paying for such training and providing release time. The administrator should determine the following: Did the individual have any training, in-service, teacher preparation, course work, or the like that relates to the focus of the communication? Principled management requires that the administrator consider whether he or she has provided the remedial assistance needed so the employee can succeed.

Likelihood of recurrence. Determination of the likelihood of recurrence of the same type of conduct seems to require a crystal ball. It is reasonable and necessary to speculate or predict into the future based on conduct that has occurred in the past. When the administrator has identified that a pattern of performance is occurring, he or she should point this out in a memorandum to the employee, using words to the following effect: "I am concerned that your conduct is likely to recur based on the fact that to date you have failed to take any steps to rectify the situation." This type of language places the determination of likelihood of recurrence in the administrator's hands. The administrator should answer the question "Is there a continuing pattern of conduct that persists despite attempts at remediation?" The principle here is that the administrator should be sure to consider the current incident in the context of overall performance and not overreact to a single incident (unless the single incident is truly flagrant or harmful to students or others).

Chilling effect on legal rights. Demands on teachers should not violate any rights they have as employees under federal, state, or other relevant laws or collective bargaining provisions. Whenever a request is made of an employee, it should be educationally sound and supported by sufficient reason, not just a preference of the administration that has no demonstrable, educational rationale. The pertinent question is "What are the rights of the individual balanced against the rights of the people who may be affected or harmed by the conduct?" Weighing the competing rights of all parties brings into play the principle of balance and justice, allowing for a wise, informed decision.

Incompetence/Unsatisfactory Performance as a Basis for Dismissal

Incompetence has been described as "a relative term which may be employed as meaning disqualification, inability or incapacity. The term can refer to lack of legal qualifications or fitness to discharge the required duty" (Gifis, 1984). Incompetence or unsatisfactory performance must be proved by a pattern of conduct rather than a single instance of behavior. Proof of charges of incompetence can come from good classroom observations, reviews of portfolios, client surveys, student outcomes, and other related evidence of performance, documented over the period of time in question, as well as thorough summary evaluations. Other evidence and testimony of students and parents can also be important.

The most important factor in proving a case of incompetence/unsatisfactory performance is the ability to show that the individual has been specifically notified of the areas that are deficient, provided a remediation plan by his or her supervisor, and given the assistance and time needed to correct these inadequacies in classroom performance (McGrath, 1993). In such cases the administrator's credibility and effectiveness are at issue. The administrator should be shown as a compassionate and helpful partner to the teacher in his or her attempts to improve teaching performance.

Other important sources of information in incompetence cases are parent and student complaints. It is essential that when a parent indicates a concern about a teacher, the administrator request that the parent put his or her concerns in writing. If the matter is resolved satisfactorily and does not form the basis of an ongoing pattern, the document probably will be of little importance. If there is an ongoing pattern of incompetent conduct, however, the document will prove to be invaluable. It will be used to refresh the recollection of the parent as to what his or her concerns were at the time. The parent complaint procedure of the school district must, of course, be followed.

A frequently asked question is "What weight is given to the complaints of students?" A single complaint by a student regarding the competence of a teacher does not carry much weight. However, an ongoing pattern of complaints received by the administration against a particular teacher will help document that there is an ongoing pattern of incompetence or unsatisfactory performance. Additional sources of information, such as tests, quizzes, transfer requests, nurse's referrals, and disciplinary referrals, will probably also be needed to show a pattern of conduct.

Remediation Plans Are Crucial to Success

Any employment action taken against an employee should be preceded by an honest effort to remediate the deficient performance, unless the offense is so egregious that it calls for immediate dismissal (Frels, Cooper, & Reagan, 1984). Remediation plans should address the following questions:

1. What specific conduct or behavior must improve?
2. What resources will be provided to address the directed change (e.g., workshops, counseling, monitoring, mentoring, retraining, written technical assistance)?
3. What measures will be used to determine if the individual has been successful in meeting the directive(s) for change?
4. Is it appropriate to give a warning that disciplinary action will be taken if the behavior does not improve?
5. How will the administrator and the employee know if the requirements have been satisfactorily met?
6. When will the administrator communicate again with the employee to check progress?
7. Has the administrator listed any pertinent attachments, such as letters from parents or copies of previous memos to the individual on the behavior/conduct/incident in focus?

What Is the Architecture of Accountability?

To understand the interrelatedness of the myriad elements that are part of employee performance and the legal analysis, it is helpful to use the metaphor of a building. To stay in the spirit of the topic, imagine an old schoolhouse. This metaphorical schoolhouse is made up of a floor, walls, and a roof (see Figure 11.1). The "architecture of accountability" can be viewed as similar to the structure of that schoolhouse. Consider the floor to be made up of the teacher's classroom and other work-related performance. The materials used to assess the quality of *performance output* consist of the following:

- *Performance standards* clearly communicated to all employees
- *Accountability records* of systematic monitoring of performance
- *Classroom observations* performed with objectivity and skill
- *Evaluation documents* that assess the whole, indicating patterns and setting forth plans where needed

These four elements are summed up in the acronym PACE.

Continuing the metaphor of the schoolhouse, the next part to consider in the architecture of accountability is the walls, which represent the *environmental impact* of the performance output. This

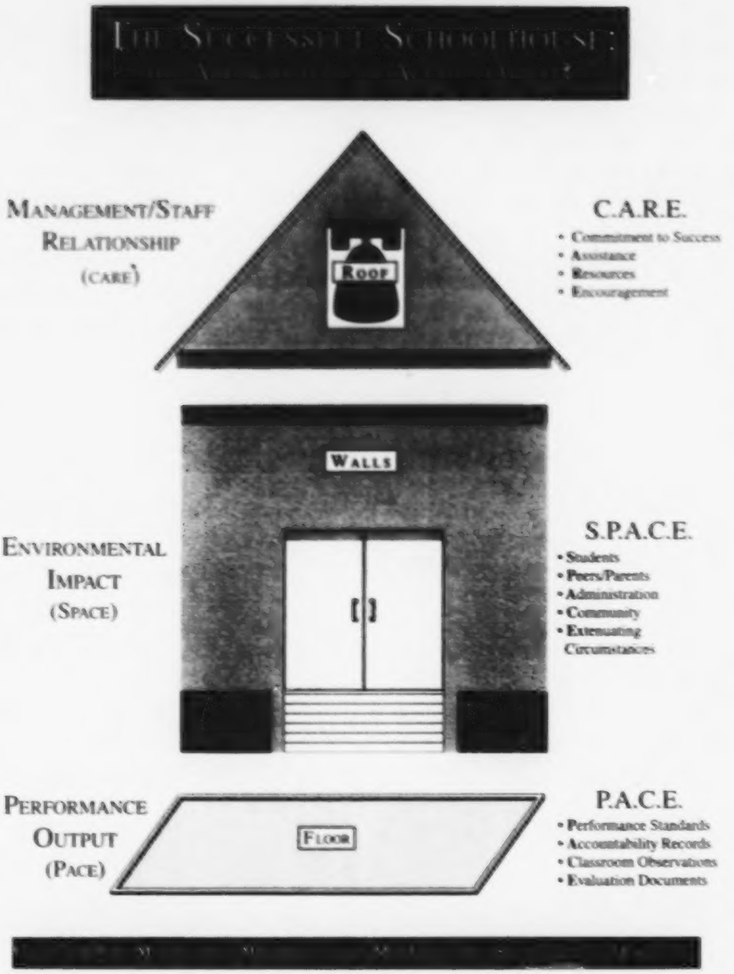

THE SUCCESSFUL SCHOOLHOUSE:
MANAGEMENT/STAFF RELATIONSHIP (CARE)
ROOF
C.A.R.E.
• Commitment to Success
• Assistance
• Resources
• Encouragement
WALLS
ENVIRONMENTAL IMPACT (SPACE)
S.P.A.C.E.
• Students
• Peers/Parents
• Administration
• Community
• Extenuating Circumstances
PERFORMANCE OUTPUT (PACE)
FLOOR
P.A.C.E.
• Performance Standards
• Accountability Records
• Classroom Observations
• Evaluation Documents

Figure 11.1. The Successful Schoolhouse: The Architecture of Accountability
SOURCE: McGrath Systems, Inc. Copyright 1996 by Mary Jo McGrath. Reprinted by permission.

part of the structure is made up of the impact of the teacher's performance as measured by the following:

BEST COPY AVAILABLE

- *Student* information, taking into account achievement and the variables that affect performance
- *Peer and parent* observations and opinions of the effect of the performance being measured
- *Administrative* factors that arise as a result of the teacher's performance, such as delayed reports or increased staff workload
- *Community relations*, as affected by the reputation the teacher and school have earned based on performance
- *Extenuating circumstances* that influence the performance

These elements are summed up in the acronym SPACE.

Finally, the roof of the schoolhouse is made up of the *management/staff relationship* and reflects the interpersonal skills of the manager as assessed by the following factors:

- *Commitment to success* of the teacher demonstrated by management
- *Assistance* provided to the teacher to accomplish remediation
- *Resources* allocated to the teacher's success
- *Encouragement* reflected in consistent feedback provided to the teacher

These elements are summarized by the acronym CARE.

If the jobs of supervision and evaluation are addressed, including all the elements represented in the architecture of accountability, administrators will succeed on many levels. They will be managers who are up-front, fair, and balanced, and will be complying with all aspects of the law of accountability. They will be both legally fit and principled.

Evaluation and Documentation: Why Are They So Important?

An administrator's opinion that a given teacher should not be in the classroom will not be accepted at face value. The burden of proof is on the school district, which must prove by a preponderance of the evidence that the teacher should be dismissed, meaning that it is 51% more likely than not that the district's charge is true. The difficulty of that task is reflected in the percentage of cases in which school districts' decisions to terminate particular employees have not been upheld (Bridges, 1992).

Regardless of the difficulty, the needed action is systematic documentation of any existing problems, which leads to remediation or substantiates the case for dismissal. This documentation needs to be created in a systematic and contemporaneous fashion, as events occur. The primary vehicles for recording those actions are (a) accountability records of incidents that occur regarding work performance and their impact; (b) classroom observations, which memorialize a teacher's performance at a particular time and place; and (c) summary evaluations, which should pull together in a periodic fashion all the factual data regarding performance (McGrath, 1993).

In a teacher termination proceeding, the evidence presented must accurately substantiate the school district's decision to proceed to dismissal. Weak, inaccurate, careless, nonexistent, or procedurally improper evaluations and documents can greatly reduce the strength of a school district's position during a termination hearing.

Administrators who are assigned to evaluate teacher performance should receive extensive training in both supervision and evaluation processes. A simple truth is that the person on "trial" in a dismissal case is not necessarily the teacher, but rather his or her immediate supervisor. Insinuations and outright attacks on the administrator are frequent, and it is essential to establish administrator competence and credibility. One good way of establishing an administrator's competence is through an assessment of the quality of documentation he or she has produced.

Careful planning and teamwork among the supervising administrator, the administrator in charge of personnel, the superintendent, the governing board, and the attorney for the school district are essential. It takes a team and a commitment of resources to equip administrative staff to do the job of supervision, evaluation, and, on rare occasions, dismissal.

Understanding the Culture and Ethics of Accountability

Punishment Is Not the Way to Alter Culture

For any significant advancement to occur in the quality of education or in the enhancement of the environment in which teachers work, a transformation must occur. We must shift from a culture in which accountability is seen as punitive to one in which it is experienced as an opportunity for full self-expression and contribution. We tend to focus on the "doingness" of what should happen to remove ineffective teachers, training administrators in how to document unsatisfactory performance and apply progressive discipline. In the meantime, we have virtually ignored the "willingness" of educators to hold themselves and others accountable for the results that indicate quality education. Additionally, we have largely ignored the culture and climate in which that willingness arises.

After a career of grappling with the issue of accountability and training administrators in effective supervision and evaluation, Bridges (1992) notes, "The challenge for scholars and practitioners alike is to find an answer to this fundamental question. 'How do we create organizational environments in which people willingly own, rather than deny or sidestep, performance problems?' " (p. 179).

Accountability Should Be an Opportunity, Not a Punishment

We can accomplish only so much when we set out to convince people to do what we want them to do, then they're going to do it their own way. School systems that mandate managers' strict supervision and evaluation of employees may experience success in the removal of unsatisfactorily performing people. However, long-term systemic change in the culture of accountability is highly questionable and may be negatively affected by a concentrated drive to rid the system of ineffective performance, especially if that drive fails to include increased communication with and acknowledgment of the successful teachers who make up the bulk of the teaching staff.

An example of such partial success is found in the case of a North Carolina school district that initiated a drive to rid its system of incompetence. The school board did a truly fine job of coordinating the efforts of administrators, legal counsel, and mentor teachers to provide an approach to identifying and evaluating less-than-satisfactory performers. Through this approach, the district identified 134 teachers who needed significant improvement and within 2½ years removed 61 of those teachers, mostly through resignations and 4 through the formal dismissal process. Another 10% of the identified teachers lifted their performance to a standard at or above acceptable performance. The district was apparently successful, except that now it reports an unexpected backlash—the "fear factor." Effective employees expressed fear that they would be "next" and sought the protection of their union from this perceived onslaught on their ranks. Although they were consulted in designing the approach and agreed with the need to remove incompetence from their ranks, when it became a reality, it struck a nerve deep within.

This kind of reaction to the imposition of accountability is a phenomenon discussed by Argyris (1991):

> Many professionals have extremely "brittle" personalities. When suddenly faced with a situation they cannot immediately handle, they tend to fall apart. . . . There is no better example of how this brittleness can disrupt an organization than performance evaluations. . . . Indeed, a poor evaluation can reverberate far beyond the particular individual involved to spark defensive reasoning throughout an entire organization. (p. 25)

Argyris attributes this brittleness to a very common fear of failure. In studies he conducted with more than 6,000 people, he determined that the defensive reasoning individuals use to place the blame for failure outside of themselves is universal (Argyris, 1994).

Bridges (1992) has noted that his own analysis of the problem of managing the incompetent teacher underestimated the significance of one variable—owning the problem: "What I did not fully appreciate was how the responses of both teachers and administrators reflected a failure to own the problem. Nor did I realize how resistant people are to owning their problems and doing something about them" (p. 177). Grove (1985) identifies five steps to problem solving: Ignore, deny, blame others, assume responsibility, and find a solution. He maintains that the process gets stuck at the "blame others" stage, and provides no insight into the organizational conditions that nurture or thwart problem ownership.

Conclusion

We must recognize that the current paradigm of accountability, which is linear rather than multidimensional and dynamic, flies in the face of people's innate understanding of the complexity of human behavior. This mismatch may result in administrators' not using such evaluation systems for the real management and assistance of their em- ployees. Rather, they may use such systems only to "do what they're told," to satisfy school district requirements of evaluation. They do their real management of people "off the record," where they can employ their recognition of the whole person in context, both personal and professional.

What if we redefined accountability as an opportunity to learn from the natural outcome of shared vision, values, and commitment to the group's goals, considering the group to be an integrated, whole, dynamic organism? What if we deeply grounded our behaviors in principles and ethics that enhanced relatedness while calling for growth and excellence? Could excellence in education become a reality for all?

Teacher dismissal is the ultimate statement of accountability. It is the final process in which a school district calls a teacher to account for his or her failure to perform satisfactorily and stands accountable for its part in the failure to remediate the situation. Perhaps a key reason it is so difficult to take the step to terminate employment is that the current paradigm holds failure as an invalidation of the person's "self" that results in the need for punishment.

Another possible paradigm includes the view that accountability represents an opportunity to be part of the whole, to matter and to make a positive difference with our life's work, and if not where we are, then somewhere else. If we are failing and unable to course correct, we are being "told" we are in the wrong place to make the contribution we want to with our lives. Argyris (1994) notes that "people genuinely strive to produce what they intend. They value acting competently. Their self-esteem is intimately tied up with behaving consistently and performing effectively" (p. 79).

This interpretation of failure takes a great deal of maturity of self and cultural evolution. It is the end product of a shift in paradigm that we are only now glimpsing. The individual being terminated from employment may never reach this view of the whole situation, but the trier of fact and decision makers must, because (a) on a personal basis they need to feel good about what they are doing, (b) they need for their actions to fit with their values and principles of how people should be treated, and (c) legally, the judicial system is designed to ensure fair treatment of public employees, procedurally and substantively.

The law of accountability is not limited in its usefulness to when a teacher is to be terminated. We can learn and incorporate the legal criteria, and their ethical and principled underpinnings, for what they are—clear guidelines as to what factors need to be in place to determine (a) whether a human being has been treated well in the workplace, and (b) what the future should be for that person based on his or her accountability for producing excellence in education.

References

Alexander, K., & Alexander, D. M. (1985). *American school law* (2nd ed.). St. Paul, MN: West.

Argyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review, 69*(3), 20-30.

Argyris, C. (1994). Good communication that blocks learning. *Harvard Business Review, 72*(4), 77-85.

Bassett Unified School District v. Commission on Professional Competence, 201 Cal. App. 3d 1444 (1988).

Bridges, E. M. (1992). *The incompetent teacher: Managerial responses*. Washington, DC: Falmer.

Cleveland Board of Education v. Loudermill, 105 S. Ct. 1487 (1985), 721 F.2d 550 (Ct. App. 1983).

Frels, K., Cooper, T. T., & Reagan, B. R. (1984). *Practical aspects of teacher evaluation*. Topeka, KS: National Organization on Legal Problems of Education.

Gifis, S. H. (1984). *Law dictionary*. Woodbury, NY: Barron's.

Grove, A. (1985). *High output management*. New York: Vintage.

Marshall, S. P. (1995). The vision, meaning and language of educational transformation. *School Administrator, 1*(52), 8-13.

McGrath, M. J. (1993). When it's time to dismiss an incompetent teacher. *School Administrator, 50*(3), 30-33.

Miller v. Chico Unified School District, 24 Cal. 3d 703 (1979).

Moore v. May Dept. Stores Co., 222 Cal. App. 3rd 836 (1990).

Morrison v. State Board of Education, 1 Cal. 3d 214 (1969).

12

Linking Teacher Evaluation and Professional Development

JOYCE ANNUNZIATA

Teacher evaluation is viewed by many as a two-edged sword: One side purports to prod professional growth, and the other side is poised above the practitioner demanding accountability for use in an employment or licensure decision. No wonder people either ignore or fear the evaluation process; no wonder evaluators either perform ritualistic observations or provide feedback that is perfunctory at best; no wonder evaluation systems vary in styles that encompass checklists, rating scales, counting or scoring positive and negative behaviors, and conferencing about or maintaining portfolios of accomplishments; no wonder practitioners search for answers to replace the inspection paradigm; no wonder state mandates and statutes regarding licensure and approval mechanisms for appraisal systems are shifting to the purview of school districts; no wonder personnel evaluation systems mentioned in reform literature remain in the stratosphere of research.

We in the school districts, at the school sites, and in the universities are trying to confront the issues in personnel evaluation. There are those of us who believe that an evaluation system not only drives performance but is the heart of professional growth for the practitioner. There are those of us who believe that an evaluation system can operationalize the standards for the profession and ensure that all teachers meet these standards to promote student learning. There are those of us who believe that an evaluation system can provide the road map for the new teacher and the avenue to follow for the advancement of the experienced teacher, and can be the vehicle for documenting the incompetent practitioner. There are those of us who believe that evaluation is the engine of the reform movement. If you are one of us, what follows may support the awareness that professionally developing all stakeholders in a teacher evaluation system is the foundation of the evaluation process. No, it is more than that—it is the bedrock of successful implementation of a teacher evaluation process embedded within a culture of professional growth for all personnel.

What the Literature Reveals About Teacher Evaluation Practices

Literature from the past 10 to 15 years emphasizes staff development as key to effective teacher evaluation practices that promote teacher growth and improvement. Zakariya (1985) cites a RAND Corporation study that delineates the following problems in teacher evaluation practices:

- Lack of resolve and competence in principals
- Resistance or apathy of teachers
- Inconsistent evaluation practices between schools
- Inadequate training for evaluators

In reflecting on the work of Scriven, Soar, Medley, and Coker, Stufflebeam and Brethower (1987) mention the following criticisms of certain personnel evaluation practices:

- Allowed unqualified persons to assume important professional roles in education
- Did not aid in ridding the education system of incompetent or unproductive personnel
- Did not provide direction for staff development

- Did not adequately recognize, and help to reward, outstanding service
- Did not provide evidence that would withstand professional and judicial scrutiny

Stufflebeam (1980) has also asserted that in order to build a more stable knowledge base in teacher evaluation, we need further work in three areas:

- Professional standards to prescribe the ethics of conduct
- Research to understand the complex structures and relationships of the discipline
- Training to ensure the professional stature of the field

When Ellett and Garland (1987) studied teacher evaluation practices in the 100 largest U.S. school districts, they found the following:

- Teacher evaluation practices had not changed since 1979.
- Local practitioners were ignorant of recent developments.
- Large districts stressed summative uses of data more frequently than formative uses.
- Few districts reported policies that included performance standards or orientation for evaluators.
- Local evaluator training was insufficient.
- Local teacher evaluation instruments included criteria that were in opposition to research literature.

A recent replication of this national study by Loup, Garland, Ellett, and Rugutt (in press) suggests that little has changed in the past 10 years. Thus, the literature indicates the need for further exploration in the field of teacher evaluation and the necessity for the supportive professional development of evaluators and practitioners to make the evaluation process do what it is supposed to do—drive teacher performance to advance practice in the classroom for student learning.

Teachers can grow when the performance criteria are clearly specified, classroom practice is regularly monitored, adherence to standards is promoted and modeled, and continuous assessment of practice and professional growth are embedded in the culture of the school and not used only to weed out incompetent teachers or because district policy or state statute requires compliance. Growth is critical to the organizational structure of the school and occurs as a direct result of relevant professional development based on identified needs of the site, community, and teacher.

How Is Professional Development Viewed Today?

No longer the stepchild of a district's operation, implemented at the whim of budgetary constraints, professional development is now welcomed into the family of reform measures for school improvement. Since *A Nation at Risk* engendered national reform efforts (National Commission on Excellence in Education, 1983), a multitude of initiatives and research-based programs have been promulgated to increase student learning. Findings from evaluations of national reform programs and restructuring movements indicate that no change in classroom practice—and thus no gain in student achievement—can be attained without intensive, ongoing quality staff development of the instructional force. The best-researched strategies and programs cannot be translated into practice unless practitioners are immersed in professional growth activities.

The newly disseminated standards for staff development from the National Staff Development Council (1995a, 1995b, 1995c) propose a radical change to the "sit and get" model delivered to participants in the past. Staff development today must provide high-quality ongoing programs with intensive follow-up and support, as well as other growth-promoting processes such as study groups, action research, and peer coaching. In addition, staff development is no longer viewed as something that is necessary only for teachers; rather, it is important for all who have an impact on student learning, including support staff and parents.

What Is New About Professional Development?

Professional development is described in the 1995 U.S. Department of Education document *Building Bridges* as the bridge that takes practitioners, the new as well as the experienced, from where they are to where they need to be to guide each student in learning. Reinforcement of the essential role professional development plays in successful education reform is found throughout this concise leaflet. The powerful notions of professional development encompass the following:

- Stressing rigorous and relevant content
- Supporting site-based initiatives
- Modeling strategies and pedagogy
- Engaging participants in ideas and materials
- Ensuring the preparation of teachers through organizational supports
- Focusing the career-long development of teachers and others who influence the learning environment
- Providing time to learn and time to practice
- Viewing professional development as an integral part of teachers' work

That Was Then; This Is Now

The current literature recognizes the problem with the image of staff development—one-shot deals, superficial and faddish programs, feel-good sessions, make-and-take or bag-of-tricks content, consultant-driven presentations, break-the-ice activities, game- and role-playing substitutes for content—which is being refocused on how learning occurs and restructured by emerging standards in the field (Gough, 1996; Sykes, 1996). These standards encourage teachers to do more than passively accept what is being delivered—to discuss, think about, try out, and hone new practices (Lieberman, 1995). The teacher as adult learner must experience what the profession expects to happen in classrooms with student learners—a wide array of learning opportunities that engage students in creating, solving real problems, and working with others (Darling-Hammond & McLaughlin, 1995; Lieberman, 1995).

Although the literature points the way to how students learn, we know the teacher is the key—not a textbook, not a computer, not a constructivist learning focus, or any other element of teaching. The teacher is the heart of every classroom every day, with every student in his or her charge. As Sykes (1996) emphasizes, student learning relies on teachers, and their professional development is vital to that learning and to any reform effort.

Teacher learning is a natural outgrowth of personnel evaluation. Thus, implementing an effective teacher evaluation system should give birth to ongoing professional development.

Professional Development: The Missing Link Between Evaluation and Improved Performance

Part of the problem with implementing a growth-promoting evaluation system is the issue of "training." This is more than just an old term; it is a concept that connotes long hours listening to trainers who bombard passive participants with rules, procedures, methods, and manuals. Training works well for inculcating repetitive actions; it does not promote continuing growth.

More often than not, training the users of an evaluation system involves a low-level, more-of-the-same presentation of the mechanics of how to handle the evaluation process. It would seem that if an evaluation system is to be a part of the culture of the school, promoting the professional growth experiences of the staff, the professional development program should be cranked up to a higher level. It must be one that revolves around sets of ideas and develops new understandings about good evaluation systems. It must suggest thoroughness through practice, consistency through follow-up, adherence through conversation, and continuity through faithfulness to stan-dards. And just who should be involved in this professional de- velopment? Professional development is for everyone involved in the evaluation system. Included in "everyone" are administrator-assessors, teacher-assessees, resource personnel who provide peer assistance, district support staff, legal department staff, and those with line responsibility to school sites. There is an ethical basis for developing all the stakeholders in the evaluation process: All must be fully apprised of the evaluation system's procedures for use, the criteria by which individuals are assessed, and the growth activities that promote improved classroom practice.

Attention must be refocused away from the 2 to 5% of the staff who need assistance and specific direction to remedy deficiencies and grow in practice. Instead, energies must turn to the majority of instructional personnel, who can and want to thrive in a school culture of professional growth. These practitioners deserve the opportunity to advance practice and be provided with other professional options that involve risk, action research, and imagination. We need to consider both groups when developing teacher evaluation training—albeit the groups are disproportionate in size. Advancement of both groups of personnel (effective and noneffective practitioners) at the school site is critical to the growth of the organization.

Interestingly, most school organizations view staff development as separate and independent from teacher evaluation (Duke, 1995). This concept is difficult to understand, given that professional development is the primary goal of most evaluation systems. What could be more obvious than to relate what teachers need to learn through professional growth activities of structured staff development programs to what is expected in a classroom, observed by an assessor, or judged as critical in attaining school improvement? To paraphrase, school improvement obtained through reform initiatives is absolutely dependent on the staff development specifically arising from what is going on in classrooms as determined by the teacher evaluation process. There is no separation or distance from one another—reform, professional development, and teacher evaluation are equal ingredients of the mix.

When Does Professional Development in an Evaluation System Stop?

The short answer to the above question is, "It doesn't." Before, during, and after the implementation of the evaluation system, assessors and all instructional personnel assessed by the evaluation system must participate in quality professional development programs that address the criteria for assessment and the procedures for using the teacher evaluation system. Professional development is necessary to certify assessors, to apprise teachers of the expectations of performance inherent in the evaluation system, and to provide district support staff with an understanding of the indicators of the evaluation system and its processes.

Vital to the ethical implementation of the evaluation system is the promotion of a sound knowledge base of the system as it relates to the context of how learning occurs. A point to remember is that assessors gain insight into the system as well as drift away from the intent of the evaluation instrument and procedures over time as they implement the evaluation process. Therefore, professional development must be ongoing and consistent throughout the use of the evaluation system, and monitoring the implementation of the teacher evaluation process is a major factor in maintaining the quality of the system.

Large districts require a cadre of specialists to develop new assessors and update previously certified evaluators. Small districts should look toward a model of internal consistency checks, where all those involved in assessment meet periodically, observe a common lesson, score the performance against the determined criteria, and compare decisions through discussion to arrive at like conclusions. What is important is that all users of the evaluation system maintain consistent implementation of the criteria and procedures and that they do this through periodic, specifically designed professional development activities.

Symbolic Training

According to Ellett and Garland (1987), training in the evaluation system appears to be inherent in most large-scale programs for both administrators and teachers. Programs for administrators have encompassed several hours to 4 or 5 days. More recently, Loup et al. (in press) have found that assessor professional development in large districts comprises as much as 2 weeks. Teachers engage in awareness sessions that include televised video programs, attend a large group meeting that presents the evaluation system to the newly hired teacher as part of an overall orientation program, receive a "mini-manual" of the criteria delineated in the teacher evaluation system, and are provided with school-site reinforcement of the assessment components and evaluation procedures.

Let us question the quality and intensity of this "training." Is it ongoing? Does it provide follow-up to the sessions to ensure understanding or change of practice without high-stakes employment decisions being imposed? Is the evaluator monitored for consistency in the implementation of the evaluation system? Are those evaluators with less-than-adequate skills in clinical supervision being afforded assistance in identifying and providing remedy to the deficiencies noted in the performance of the teacher? Is adherence to the criteria for assessing performance being interpreted the same way? If the training does not confront these concerns, then it is merely symbolic—a satisfaction of the "notice" requirement set forth in statute or rule—a superficial compliance with the NEAT (notification, explanation, assistance, and timeline) process, but not professional development.

The NEAT process is formative in nature: the supervisor *n*otifies the teacher of the standard or criteria to which the teacher must adhere; the supervisor *e*xplains what is expected of the employee in the way of performance and adherence; the supervisor *a*ssists the teacher in attaining the expected performance; and the supervisor provides a *t*imeline by which the performance is to comport with the expectation. When applying the NEAT process to teacher evaluation, the site administrator must notify the teacher of the assessment criteria prior to observation of performance; subsequent to the observation, the observer should delineate strengths and explain the weaknesses in the performance; during this explanation, the administrator should provide assistance by developing suggestions for improvement during a specified time.

This feedback to the teacher by the assessor is considered a most important structure and probably the most direct way to assist teachers in the ways they teach and the ways students learn (Nevo, 1994). According to Scriven (1967), the formative use of evaluation is to provide feedback to the teacher for improvement, whereas the summative use of evaluation is for selection and accountability of instructional personnel. Thus, this formative use of the evaluation system is an example of professional development both for the teacher, who is enlightened as to what is going on in the classroom through another's perspective, and for the observer, who is articulating supervisory skills and honing professional growth strategies.

If training is to be more than symbolic, the following key components must be understood and implemented:

- Professional development is fundamental to the implementation of the evaluation system.

- Knowledge about the personnel evaluation system and its procedures must be internalized by the users as well as the recipients of the evaluation system.
- Professional development is an occasion of growth for both the evaluator and the teacher.
- The legal challenge is a test of the evaluator's skills in communicating the indicators of teaching and learning.

Who Should Deliver the Professional Development Program for Administrators, Teachers, and District Staff?

My personal experience in large district evaluation system professional development indicates that the developer(s) of the system should be integrally involved in the planning and delivery of the training programs. This is especially important during the "training of trainers" for the district to ensure consistency of the program. Supporting the developer, spokespersons from the teacher labor union or professional association are critical to the process, especially with the teachers who are introduced to the evaluation system and, during appropriate intervals, to all teachers in order to update and sharpen awareness of the requisite components of the system.

Small school districts must be as aware of consistency of training programs as larger districts. In my own experience I have found that collaboration between the developer and the person in charge of professional development is vital during the initial stages of implementation. Additionally, this association is integral to the planning and delivery of an effective training program.

In addition to traditional delivery systems, today's technology lends itself very well to conveying the "message" of teacher evaluation to all teachers at any time. Videos, computer disks, distance learning techniques, and collaborative partnerships with local colleges and universities can assist in the dissemination of information to teachers concerning what is required to be done in a classroom and how to do it. Traveling to a meeting is no longer necessary, nor is the presence of a trainer or cadre of consultants. Interaction can be encouraged by use of telephone, self-help CD-ROMs, and collegial study groups. "Virtual" training sessions can be held wherever a teacher or team has access to a television or computer.

What Is Effective Professional Development for Implementation of the Evaluation System?

Model teacher evaluation professional development programs provide practitioners with simulated classroom experiences to assess through videotapes or role-play. The opportunity to visit actual classrooms to try out the evaluation system is encouraged. Additionally, multiple data sources are used to certify assessors. To ascertain evaluator proficiency in the use of the evaluation system, professional activities might include written tests, videotaped classroom performance, rating scales, self-assessment instruments, structured interviews, and questionnaires.

Professional Development Is the Linchpin Between Performance and Growth for the Administrator and Teacher Practitioner

The focus of the teacher evaluation system must never waiver from the main goal of education—student learning. The purpose of the evaluation system is to engender growth in the instructional personnel being assessed, growth that will affect student learning. Formative assistance is provided to instructional staff prior to implementation of the system at a given site. Additionally, feedback by the observer to the teacher after observation is a valuable tool for professional growth.

Without a commitment to ongoing professional development on the part of administrators and teachers, the best evaluation system will not work. For professional development to be effective, the following elements should be addressed:

- A commitment from the top—the superintendent and the school board—is vital to the successful implementation of the evaluation system.
- A commitment from the teacher labor union or professional association is integral to the development of teachers and necessary for the requisite revisions and modifications to the system.
- Time spent in developing and delivering effective professional development programs will save time as the teacher evaluation system is used in the field.

- Monitoring of the implementation of the evaluation system is vital to the system's effectiveness. Responsibility for oversight of the teacher evaluation process from the initial development of stakeholders through the final throes of litigation should be determined and designated.
- No one should embark upon documenting for incompetence unless he or she is committed to the task. An administrator's role as evaluator will diminish the trust relationship (maternalistic or paternalistic perception) that exists between the supervisor and teacher. Administrator documentation will be subject to scrutiny, grievance procedures, and legal challenges. In other words, the faint of heart need not apply.

Summary

The following are some ways in which administrators can help make meaningful connections between personnel evaluation and professional development:

- Focus practitioner attention on the benefits of teacher evaluation as it promotes growth in the profession.
- Spend more time building a culture of receptivity of professional development and change.
- Forgo the emphasis on paper-driven documentation exercises and promote professional conversations.
- Encourage practitioners to engage in learning and growing so as to promote the same in their students.
- Return to the drawing board and reengineer the personnel evaluation system by searching for answers to new concerns.

In conclusion, I want to emphasize the following points:

- Teacher evaluation systems must be carefully developed to be formative in focus and growth producing.
- Traditional approaches to training have led to a mind-set that is anti-professional development.
- Professional development must lead to common understandings and common beliefs.

- Joint professional activities between teachers and administrators, in alignment with norms and the culture of the school, are important to the improvement of student learning.

A Final Note

In this chapter I have integrated the literature on professional development and teacher evaluation with my own personal experiences gained from 15 years as a trainer in a district's teacher evaluation system and as an administrator in the office that monitors the performance of all employees. Coloring these views of personnel evaluation research and practice also is my extensive experience in litigation as an expert witness in performance appraisal systems and personnel standards. It is my hope that this chapter will evoke a shared memory of where we as practitioners in teacher evaluation have been and contribute to a vision of where we want to go—and the route is professional development.

References

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76,* 597-604.

Duke, D. L. (1995). *Teacher evaluation policy: From accountability to professional development.* Albany: State University of New York Press.

Ellett, C. D., & Garland J. S. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to "state-of-the-art" systems? *Journal of Personnel Evaluation in Education, 1,* 69-92.

Gough, P. B. (1996). No more "pie in the sky"! *Phi Delta Kappan, 77,* 459.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *Phi Delta Kappan, 76,* 591-596.

Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (in press). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education.*

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: Government Printing Office.

National Staff Development Council, in Cooperation with the National Association of Elementary School Principals. (1995a). *Standards for staff development* (Elementary school ed.). Oxford, OH: Author.

National Staff Development Council, in Cooperation with the National Association of Secondary School Principals. (1995b). *Standards for staff development* (High school ed.). Oxford, OH: Author.

National Staff Development Council. (1995c). *Standards for staff development* (Middle school ed.). Oxford, OH: Author.

Nevo, D. (1994). How can teachers benefit from teacher evaluation? *Journal of Personnel Evaluation in Education, 8,* 109-117.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

Stufflebeam, D. L. (1980, May). *Standards, research, and training: Three priorities for professionalizing educational evaluation.* Paper presented at the Minnesota Evaluation Conference.

Stufflebeam, D. L., & Brethower, D. M. (1987). Improving personnel evaluations through professional standards. *Journal of Personnel Evaluation in Education, 1,* 125-155.

Sykes, G. (1996). Reform of and as professional development. *Phi Delta Kappan, 77,* 464-467.

U.S. Department of Education. (1996). *Building bridges: The mission and principles of professional development.* Washington, DC: Government Printing Office.

Zakariya, S. B. (1985, July). These are the elements of a sound teacher evaluation system. *American School Board Journal, 172,* 25.

Author Index

Abrami, P. C., 156, 157
Adams, F. E., 66, 69, 74, 81, 84, 86
Airasian, P. W., viii, xii, xiii, 56, 178, 189, 215-217, 219, 220, 222, 224-227, 229, 242, 245
Aleamoni, L., 143, 145, 146, 157
Alexander, D. M., 67, 86, 273, 287
Alexander Hamilton Institute, 256, 260, 261, 263, 266
Alexander, K., 67, 86, 273, 287
Alkin, M. C., 61, 62, 74, 86
Allen, D. W., 225, 242
American Association of School Personnel Administrators, 31, 55
Anderson, G. C., 253, 258, 260, 266
Anderson, J. B., 238, 242
Angelo, T. A., 225, 227, 230, 242
Annunziata, J., viii, xii, xiii, 288
Argyris, C., 285, 286, 287
Armentrout, B. W., 268
Ashton, P. T., 239, 242
Asoko, H., 218, 243
Astuto, T. A., 2, 19
Austin, N. K., 268

Bailey, E., 157
Bailey, G. D., 224, 225
Baker, J., Jr., 268
Ball, D. M., 137, 142, 158, 159
Bandura, A., 239, 242
Barber, L. W., 216, 223, 226, 227, 242
Barger, S. A., 223, 246
Barnett, J. G., 253, 258, 260, 266
Barry, K., 56
Bartlett, E., 193, 214
Bayless, D. L., 134, 157
Becher, R. M., 135, 157
Beckham, J. C., 61, 66, 70, 72, 75, 77, 79, 82, 84, 86
Beecher, D. E., 109, 126

Bembry, K. L., 168, 189
Bennett, A. C., 7, 14, 21
Berk, R. A., 35, 55, 180, 187, 189
Berliner, D. C., 119, 126, 217, 243, 244
Bernstein, H. T., 29, 58, 60, 90, 132, 161
Biddle, B. J., 110, 126
Black, H. C., 68, 74, 76, 79, 86, 89
Boeschen, D. A., 268
Bolton, D. L., 14, 19
Bonstingl, J. J., 134, 135, 158
Borko, H., 193, 213
Boyer, E. C., 110, 128
Brandenburg, D. C., 35, 55
Brandt, R. S., 127, 151, 158
Braskamp, L. A., 35, 55
Bray, S. W., 156, 161
Brethower, D. M., 10, 23, 289, 301
Bridges, E. M., 16, 19, 79, 82, 86, 133, 158, 283, 284, 285, 287
Bringhurst, K., 244
Brophy, J. E., 32, 55, 110, 112, 126, 128
Browning, M., 146, 158
Bruner, J., 218, 243
Burns, M., 244
Bushardt, S. C., 268
Buzzotta, V. R., 252, 256, 266
Byrd, P. C., 268

Cain, B. L., 59, 60, 69, 84, 86
Calderhead, J., 219, 243
Calfee, R. C., 244
Cambron-McCabe, N. H., 70, 88
Campbell, G., 193, 214
Carnegie Task Force on Teaching as a Profession, 243
Carr, E. G., 117, 127
Carter, K., 218, 243
Castetter, W. B., 5, 14, 19
Centra, J. A., 35, 55
Chadbourne, R., 57
Charters, W. W., 108, 126
Chauvin, S. W., 118, 126
Cizek, G. J., 53, 54, 56, 177, 189
Clandinin, D. J., 218, 243
Clark, C. M., 34, 56
Clark, D. L., 2, 19,
Cochran-Smith, M., 222, 243
Cohen, D. K., 187, 189
Cohn, M. M., 225, 245
Coker, H., 16, 21, 34, 56, 93, 103, 110, 127, 130, 160, 289
Coley, D., 157
Conley, D. T., 10, 15, 19
Connellan, T. K., 14, 19
Connelly, F. M., 218, 243
Cooley, V., 244
Cooper, T. T., 63, 87, 279, 287
Corcoran, T., 60, 87
Cornett, J. W., 221, 245, 246
Costa, A. L., 34, 56, 134, 158
Court, D., 221, 243
Crawford, G., 146
Crone, L. J., 60, 87
Cross, K. P., 225, 227, 230, 242
Cullen, K., 56
Cummings, I. L., 8, 19
Cunningham, G., 178, 192

d'Apollonia, S., 156, 157
Darling-Hammond, L., 19, 29, 58, 60, 71, 87-88, 90, 93, 103, 112-113, 116, 126-128, 132, 156, 158, 161, 217, 243, 292, 300

Dawson, V., 117, 128
Day, C., 244
Delandshire, G., 218, 243
Delon, F. G., 65, 71, 87
Deming, W. E., 134, 158
Denicolo, P., 244
Denton, D. K., 268
Derven, M., 252, 253, 266
Dixon, R. G. D., 134, 158
Downs, T. M., 252, 259, 267
Doyle, W., 218, 243
Driscoll, A., 146, 158
Driver, R., 218, 243
Duke, D. L., 3, 6, 19, 225, 244, 294, 300
Dunkin, M. J., 110, 126
Dwyer, C. A., 216, 217, 244

Ebmeier, H., 146, 158
Educational Research Service, 10, 62, 76, 82, 87, 130, 135, 158
Egelson, P., 219, 226, 232, 236, 246
Elbaz, F., 221, 223, 244
Elias, P., 32, 56
Ellett, C. D., vi, xii, xiii, 15, 20, 107, 110-113, 116, 118, 126-127, 134, 158, 290, 295, 300
Elmore, R. F., 88
Epstein, J., 136, 137, 156-158
Eraut, M., 220, 244
Erickson, G. L., 239, 244, 246
Evertson, C. M., 131, 158

Farnham-Diggory, S., 221, 244
Faucette, S. L., 137, 142, 143, 158, 159
Feiman-Nemser, S., 223, 244
Fessler, R., 244
Flanders, N. A., 110, 127
Floden, R. E., 244
Follman, J., 143, 145, 146, 159
Ford, B., 217, 244
Franklin, J., 157
Frase, L. E., 81, 87, 130, 131, 132, 133, 159
Fraser, B., 117, 124, 128
Freiberg, H. J., 238, 242
Frels, K., 16, 20, 63, 64, 66, 71, 72, 74-80, 82, 87, 279, 287
Fuhr, D. L., 60, 87
Fuhrman, S. H., 62, 88, 247
Fullan, M. G., 2, 3, 20, 134, 159, 217, 245, 246
Fuller, F. F., 225, 244
Furst, N., 110, 128

Gadamer, H., 218, 244
Gage, N., 110, 127, 128
Gale, J., 247
Garland, J. S., 113, 127, 134, 158, 290, 295, 300
Garmston, R. J., 34, 56
Gates, A., 268
George, P. S., 9, 20
Gifis, S. H., 278, 287
Gilliss, G., 221, 244
Girod, 167, 170
Gitlin, A., 216, 218, 244
Glatthorn, A., 11, 15, 20
Glickman, C. D., 217, 244
Goertz, M., 60, 87
Good, T. L., 32, 55, 109, 127
Goodale, J. G., 14, 20, 268
Goodlad, J. I., 129, 159
Gore, J., 223, 244

Gough, P. B., 292, 300
Gower, R., 226, 246
Grimmett, P. P., 217, 239, 244, 246
Gronlund, N., 177, 178, 180, 189
Grove, A., 287
Guba, E. G., 117, 127
Gullickson, A., viii, xii, xiii, 56, 215-216, 219-220, 222, 224, 226-227, 229, 242
Gunne, G., 156, 160

Haak, R. A., 146, 159
Haefele, D., 130, 132, 133, 151, 159
Haertel, E., 35, 46, 51, 56, 162, 176, 180, 189
Haertel, G. D., 56, 58, 60, 90
Hagen, E., 178, 192
Hagerty, P., 212, 214
Halkes, R., 221, 245
Hanson, J. R., 225, 247
Hargreaves, A., 215, 217, 238, 245, 246
Harris, B. M., 11, 15, 20, 129, 133, 143, 146, 156, 159
Harris, W. U., 63, 88
Hartman, D., 193, 214
Hatton, N., 216, 220, 223, 236, 245
Haysom, J., 226, 245
Helm, V. M., vii, xii, xiii, 4, 5, 7, 11, 12, 15, 20, 23, 62, 71, 77-79, 84, 89,
131-132, 149, 154-156, 161, 198, 214, 251, 256, 268
Hoden, 223
Holladay, R., 157
Holley, F. M., 131, 158
Horn, 163, 167, 168, 169, 170, 178, 188, 189
Horton, J. L., 16, 20, 64, 78, 80, 82, 87
Howard, K. C., 156, 161
Hughes, C. S., 127
Hunter, M., 17, 20

Ilgen, D. R., 268
Ingvarson, L., 57
Iwanicki, E. F., 17, 20, 60, 71, 88, 226, 227, 245

Jenkins, J. M., 268
Jenkins, R., 146
Johnson, S. M., 70, 88
Joint Committee on Standards for Educational Evaluation, 10, 15, 20, 29, 56, 84, 88, 91-93, 100, 103
Jones, A., 225, 245
Jones, B. F., 117, 127
Joyce, B., 33, 56, 134

Kagan, D. M., 222, 245
Kagan, N., 225, 245
Kauchak, D., 133, 143, 145, 146, 160
Keegan, J. J., 14, 20
Kindred, K. P., vi, xii, xiii, 59
Kirkpatrick, D. L., 268
Kirp, D. L., 64, 90
Kleiber, D., 146, 159
Klein, H. J., 260, 267
Kleinsasser, A. M., 221
Kottkamp, R. B., 216, 219-220, 224-227, 238, 245-246
Krayer, K. J., 253, 267
Krein, T. J., 258, 267

Kremer-Hayon, L., 216, 221-222, 225, 227, 236, 244-245
Kuhn, D., 218, 219, 223, 245

Laing, S. O., 60, 88
Lambert, L., 34, 56
Larson, R., 146, 159
Latting, J. K., 268
Leach, J., 218, 243
Levin, B., 64, 90
Lichtenstein, G., vii, xii, xiii, 193, 196, 214
Lieberman, A., 292, 300
Linn, R., 178, 180, 189
Little, J. W., 3, 17, 20
Livingston, S. A., 53, 56
Locke, E. A., 14, 21
Louden, W., 220, 236, 245, 246
Loup, K. S., 113, 118, 127, 290, 295, 300
Loysk, B., 253, 267
Lytle, S. L., 222, 243

MacLean, J. C., 268
Mager, G. M., 135, 159
Malen, B., 247
Manatt, R. P., 15, 21
Mancision, J., 253, 267
Manning, B. A., 225, 244
March, J. G., 3, 21
Marien, B. A., 268
Mark, M. M., 136, 159
Marshall, S. P., 272, 287
Martin, B. A., 268
Marzano, R. J., 117, 127
Massaro, G., 157
McBey, K., 268
McCarthy, M. M., 70, 88
McCloskey, W., 219, 226, 232, 236, 246
McConney, Andrew, vii, xii, xiii, 36, 39, 44, 56, 162, 167, 182-183, 185, 189-190
McCutcheon, G., 221, 245, 246
McDonald, D., 157
McDonald, F. J., 32, 56
McEachern, L., 226, 227, 245
McGaghie, W. C., 4, 21
McGrath, M. J., vii, xii, xiii, 60, 61, 78, 88, 270, 279, 281, 283, 287
McGreal, T. L., 4, 11, 15, 17, 21, 130-131, 133-134, 146, 149, 157, 159, 252, 267
McLaughlin, M. W., 3, 21, 29, 58, 60, 63, 88, 90, 113, 116, 127, 132, 161, 292, 300
McNamara, D., 223, 246
McNeil, J. D., 143, 146, 160
McRobbie, C., 124
Medley, D. M., 16, 21, 34, 56, 93, 103, 108, 110, 113, 127, 128, 130, 160, 289
Mendro, R. L., 168, 189
Meyer, 252, 253, 256, 267
Michaelec, P., 193, 213
Miller, P., 156, 160
Millman, J., 19, 35, 56, 61, 71, 87, 88, 112-113, 127-128, 157, 158, 160, 162, 176-177, 180, 187-189, 243, 252, 267
Milner, J. O., 217, 246
Mitchell, A., 142, 160
Mitzel, H., 110, 128
Mohrmon, A. M., 252, 267
Mortimer, E., 218, 243
Mulkeen, T., 131, 161

Mulryan, C., 109, 127
Murnane, R. J., 187, 189
Myers, B., 244

Nais, J., 239, 246
Nathan, B. R., 252, 267
National Board for Professional Teaching Standards, 214
National Commission on Excellence in Education, 63, 88, 291, 301
National Council of Teachers of Mathematics, 31, 56
National Staff Devlopment Council, 291, 301
Needels, M., 110, 127
Nevo, D., 296, 301
Newman, R. J., 269
Niles, J. A., 223, 247
Nisbett, R. E., 223, 246
Northwest Regional Educational Library, 129, 143, 160
Nowak, J. E., 64, 65, 89
Nutt, H., 109, 128

O'Day, J., 217, 247
Ogle, D. S., 117, 127
Olsen, L. D., 7, 14, 21
Olson, J. K., 221, 245
Ornstein, A. C., 31, 57
Orsak, T. H., 168, 189
Ory, J. C., 35, 55
Osterman, K. F., 216, 219, 220, 224, 226, 227, 238, 246
Ostrander, L. P., vii, xii, xiii, 129, 137, 142, 147, 156-160
Overly, D., 133, 160
Owens, R. G., 9, 21
Ozga, J., 246

Palinscar, A. S., 117, 127
Parkinson, J., 59, 61, 66, 69, 73, 74, 89
Patten, T. H., Jr., 268
Patton, M. Q., 14, 21
Pease, P., 193, 214
Pease, S. R., 93, 103, 116, 126, 156, 158
Peck, R., 146, 159
Pennock, 253, 256, 267
Peterson, K. D., 79, 84, 89, 131, 133-134, 137, 142-143, 145-146, 154, 156-158, 160
Peterson, P. L., 34, 56, 223, 246
Peterson, R. B., 268
Petrie, B. I., Jr., 68, 74, 76, 79, 89
Petrosky, A. R., 218, 243
Pfeifer, R. S., 3, 21, 63, 88, 113, 127
Phi Delta Kappa National Study Committee on Evaluation, 14, 15, 21
Plachy, R. J., 269
Plachy, S. J., 269
Pogrow, S., 2, 21
Pollard, A., 225, 246
Pope, M., 244
Popham, W. J., 56-57, 143, 146, 156, 160-161, 189, 252, 267
Porter, A. C., 110, 128
Presseisen, B. Z., 127
Price, K., 244
Provenzo, E. F., 225, 245

Rachor, R., 177, 189
Rankin, S. C., 127
Rausch, E., 269
Reagan, B. R., 63, 87, 279, 287
Redfern, G. B., 9, 14, 21, 81, 89
Rivera, O., 156, 160
Robinson, G. E., 35, 57
Root, D., 133, 160
Rosenfeld, M., 30, 57
Rosenshine, B., 110, 128
Ross, E. W., 221, 245, 246
Ross, L., 223, 246
Rossow, L. F., 59, 61, 66, 69, 73, 74, 89
Rostad, O.P., 217, 244
Roth, Michael, 124
Rotunda, R. D., 64, 65, 89
Ruckle, C., 193, 214
Rugutt, J. K., 113, 127, 290, 300
Russell, R., 244
Ryan, K., 225, 242

Sanders, J. R., vi, xii, xvi, 10, 23, 91, 93, 103
Sanders, W. L., 163, 167-170, 178, 188, 189
Sandler, L., 269
Saphier, J., 6, 22, 226, 246
Schalock, H. D., vii, xii, xvi, 162, 166, 170, 187, 189
Schalock, M. D., vii, xii, xvi, 162, 167, 170, 189
Schön, D. A., 217, 218, 219, 220, 246
Schwab, D. P., 8, 19
Schwab, J., 221, 247
Sclan, E. M., 63, 75, 89
Scott, P., 218, 243
Scriven, M., v, xi, xvi, 7, 14-16, 18-19, 22, 27, 31-33, 37-38, 57, 58, 61, 89, 114, 116, 128, 130-132, 156-157, 160, 162, 176, 185, 189, 252, 267, 289, 296, 301
Sergiovanni, T. J., 216, 247
Shapiro, B. C., 60, 89
Shinkfield, A. J., 113, 128, 218, 247
Shotland, R. L., 136, 159
Shuell, T. J., 32, 57
Shulman, L. S., 50, 57, 119, 128, 182, 189, 194, 214, 224, 247
Silver, H. F., 225, 247
Simon, A., 3, 21, 110, 128
Skurnik, L. S., 30, 57
Smith, D., 216, 220, 223, 236, 245
Smith, M. S., 217, 247
Smylie, M. A., 162, 187, 189
Snell, S. A., 260, 267
Soar, R. S., 16, 21, 34, 56, 93, 103, 110, 127, 130, 160, 289
Somerick, N. M., 253, 267
Stake, R. E., 89
Stallings, J., 110, 128
Stanley, S. J., 56, 57, 156, 161, 267
Starratt, R. J., 216, 247
Steffe, L. P., 247
Stevens, D., 146, 158, 160
Stevenson, C., vii, xii, xvii, 193
Stiggens, R., 181, 189, 225, 244
Streshly, W., 130, 131, 132, 133, 159
Stronge, J. H., v, vii, xii, xvii, 1, 3-5, 7, 9, 12-13, 22-23, 62, 66-67, 71, 75, 77-79, 84, 89, 129, 131-132, 149, 154-156, 161, 176-177, 189, 198, 214, 256, 268

Stufflebeam, D. L., 5, 10, 17, 23, 56, 113, 128, 186, 189, 216-218, 244, 247, 289, 290, 301
Suhor, C., 127
Sykes, G., 162, 180, 189, 292, 301

Tann, S., 225, 246
Taylor, F., 109, 128
Taylor, P., 117, 124, 128
Teacher Evaluation Task Force, Dallas Public Schools, 191, 192
Teaching Research Division, Western Oregon State College, 192
Teddlie, C., 60, 87
Tetenbaum, T., 131, 161
Theall, M., 157
Thorndike, R. M., 178, 189
Thornton, R. F., 30, 57
Tiess, P., 244
Tigges, J. H., 59, 82, 90
Timmons, M., 193, 213
Tobin, K., 117, 124, 128
Tom, A. R., 218, 247
Toma, C., 218, 247
Travers, R. M. W., 128, 160, 224, 247
Tucker, N. A., 156, 161
Tucker, P. D., vi, xii, xvii, 4, 13, 23, 59, 66-67, 75, 89, 132, 161
Tyler, Ralph W., 5, 23

U. S. Department of Eduation, 291, 301
Umiker, W., 269

Valentine, J. W., 14, 16, 23
Valli, L., 220, 223, 247
Van Manen, M., 220, 247
Vicinanza, N., 56
Vonk, H. C., 244

Walberg, H. J., 57, 60, 90
Wang, M. C., 60, 90
Waples, D., 108, 126
Waxman, H. C., 57
Webb, R. B., 239, 242, 258, 268
Webster, 162, 167, 168, 169, 177, 189
Weerasinghe, D., 168, 189
Weil, M., 33, 56
Wertsch, J. V., 218, 247
Wheeler, P. H., v, xi, xviii, 27, 29, 35, 37, 46, 49-51, 56-58, 174-176, 178, 180, 189
Whinery, B., 212, 214
Wiersma, W., 56
Wildman, T. M., 223, 247
Wise, A. E., 29, 58, 60-61, 70, 90, 93, 103, 116, 126, 132, 156-158, 161
Wittrock, M. C., 56, 244
Wolf, K., vii, xii, xviii, 174, 18, 193, 194, 212, 214
Wood, C., 132, 161

Yudof, M. G., 64, 90

Zakariya, S. B., 289, 301
Zeichner, K., 223, 244
Zieky, M. J., 53, 56

Subject Index

Accountability, 3, 61, 134, 157, 164, 181, 216, 220, 270, 282-286, 288

Center for Research on Educational Accountability and Teacher Evaluation (CREATE), 183
Classroom-based assessments, 107-128
  history, 108-112
  issues, 111
  minimum competency testing, 111
  on the job teacher evaluation procedures, 111
  problems with, 113-114
  process-product research paradigm, 110
  suggested solutions, 115-118
Client feedback, 11, 130, 135, 203, 226, 232, 278
Climate, classroom, 131
Communication, 7, 8, 18, 142, 187
Constitutional law:
  due process, 63, 64, 273
  equal protection clause, 65
  Fourteenth Amendment, 63-66
  liberty interest, 64
  procedural due process, 63-64
  property right, 64
  substantive due process, 63
  Tenth Amendment, 66

Duty-based teacher evaluation, 37-38
  assessment methods, 46, 49, 50
  descriptors, 47-49
  indicators, 46-49
  job description, 39-45
  standards, 46, 51

- Educational reform, 131, 134, 291
  - collaboration, 134
  - collegiality, 134
  - stakeholder empowerment, 134
  - teacher accountability, 134
- Evaluation conference, 251-269
  - characteristics of an effective, 258
  - classroom observation, 251
  - preparation, 253
  - purposes, 252
- Evaluation models, 12, 13, 118-121
  - beginning teacher evaluation study, 31
  - centralized monitoring and evaluation systems, 168-169, 186
  - collegial model of assessment, 149-151
  - Dallas, Texas public schools teacher evaluation system, 166-169
  - decentralized (school-based) monitoring and evaluation systems, 170-186
  - developmental phase, 13
  - duty-based teacher evaluation model, 37-38
  - expanded clinical evaluation model, 121-124
  - goals and roles model, 12-13, 17
  - implementation phase, 15
  - STAR (System for Teaching and Learning Assessment and Review), 118-121
  - teacher's work sample methodology, 170, 174-175
  - Tennessee value-added assessment system, 167-170
- Evaluation summary, 77
  - documentation, 78
  - notice, 79
  - recommendations, 80
  - time frame, 81
- Evaluation systems:
  - examples, 113
  - formative, 61
  - improvement oriented, 4
  - job analysis, 30
  - outcome oriented, 4
  - style-based, 31
  - summative, 35, 61

- Foundations of evaluation, 27, 29
  - effective teaching, 33
  - job analysis, 30
  - process-product studies, 33
  - style-based evaluation, 31
  - systematic process, 16

- Joint Committee on Standards for Educational Evaluation, 92

- Legal principles, 85
  - documentation, 85
  - equitable treatment, 85
  - job-related criteria, 85
  - legal compatibility, 85
  - opportunity to improve, 85
  - pattern of performance, 85
  - prior notification, 85
  - substantiation, 85
  - trained evaluators, 85

written policy, 85
Legal sources, 60
judicial decisions, 63
state statutory law, 63, 67
U. S. Constitution, 63

Minimum competency testing, 111
Multiple data sources, 10-12, 130, 156-157, 185, 187, 213, 298
artifact collection, 62
assistant principals, 135
checklists, 62, 226, 288
client feedback, 11, 130, 135, 203, 226, 232, 278
client surveys, 145, 157, 208, 278
competency tests, 61
department chairs, 135
external clients, 135
internal clients, 135
media recording, 226
minute surveys, 230-231
observation, 11, 62, 130-133, 203, 251, 278
parents, 135
parent survey, 143, 147
peer feedback, 147, 153
peers, 135
peer survey, 152-153
portfolios, 11, 175, 193-214, 217, 226, 278, 288
principals, 135
questioning, 62
questionnaires, 226
ratings, 132, 145, 199, 203
rating scales, 61, 74-75, 226, 288
standardized test results, 74
self-evaluation, 11, 135, 157, 215-247
self-selection, 62
student performance, 11,62, 203, 278
students, 135
student survey, 147
supervisors, 135

National Teacher Examination, 65, 111
Negotiated agreements, 69
conditions of employment, 69
procedures, 75
recommended components, 70-75
Nonproductive teachers, 270-287
dismissal, 272
evaluation documentation, 283
fitness for service, 275
incompetence, 278
just cause, in review of, 274
remediation plans, 279
unsatisfactory performance, 278

Outcomes, 35
management by objective, 35
student outcomes, 35
student ratings, 35
teacher productivity, 35

Parent survey, 143, 147
classroom environment, 142
communication, 142
grading, 142

- homework, 142
- instruction, 142
- interpersonal relationships, 142
- Lenawee Intermediate School District Parent Survey, 144-145
- parent perception survey, 137-142

Peer feedback, 147, 153-154
- guidelines, 154

Peer survey, 152-154
- Lenawee Intermediate School District Teacher Survey (peer survey), 152-153

Performance standards, 15
- job performance, 15

Personnel evaluation standards, 10, 91, 92
- accuracy, 98-100
- background, 92
- development, 93-94
- feasibility, 97, 100
- implications, 95
- propriety, 95-96, 100
- suggestions for use, 101
- utility, 96-97, 100, 198

Portfolios, 11, 175, 193-214, 217, 226, 278
- characteristics, 194
- client survey, 208
- content evaluation, 202-203
- criteria for evaluating, 198
- design, 200-201
- evaluation procedures, 211
- guidelines, 210
- performance standards, 199
- professional content standards for teachers, 200
- reliability, 198
- specifications, 201-202
- utility, 198
- validity, 198

Principles of evaluation:
- accountability, 3, 61, 134, 157, 164, 216, 220, 270, 282-286, 288
- balance, 183, 185
- classroom climate and, 9, 131
- collaboration, 134, 149, 184, 187, 216
- communication, 7, 8, 18, 142, 187
- effective, 9, 17
- goals, 6, 7
- informed decision making, 187
- improvement, 187
- multiple data sources, 10-12, 130-156-157, 185, 187
- professional growth and development, 216, 226, 239, 288, 291-297
- school improvement, 294
- stakeholders, 185, 187, 288, 293, 299
- system design, 183
- technically sound, 9, 10

Professional development, 288-301

Program evaluation standards, 92

Purposes of evaluation, 28-29, 60-61, 176
- accountability, 3, 61, 134, 157, 164, 181, 216, 220, 270, 282-286, 288
- evaluation systems, 4, 28
- formative evaluation, 176, 181
- improvement oriented, 5, 61
- individual goals, 176

- measures of student performance, 176
- organizational goals, 176
- outcome oriented, 5
- performance improvement, 3, 16
- staff development, 289, 291-292, 294
- summative evaluation, 176, 181

Restructuring, 131

Roles and responsibilities, 17, 36, 43
- credentialing responsibilities, 36
- licensure, 36

Self-evaluation, 215-247
- approach, 225
- professional development, 226
- purposes of, 217-218
- reflection, 220-222, 224, 228-229, 236
- self-reflection tools, 226-227, 230-231, 233-237
- teacher evaluation tool kit, 229

Standards:
- criteria, 46
- developmental, 52
- minimum, 53
- personnel evaluation, 10
- setting, 46, 51

State statutory law, 63, 67
- Children First Act in Louisiana, 67
- continuing contracts, 67
- probationary contracts, 67

Student feedback, 145, 232
- ability to assess, 145
- client survey, 145
- ratings, 145
- student survey, 147
- Williamsburg/James City County Public Schools Student Survey: k-3, 148
- Williamsburg/James City County Public Schools Student Survey, 9-12, 150-151

Student learning, 162
- influences on, 164
- measurements of, 162

Student performance, 172, 226
- balance of, 183
- criteria for, 178, 181
- data, 233
- measures of, 172, 176-177, 182
- uses of, 173

Teacher dismissal, 83, 272
- grounds, 83

Teacher evaluation systems, 110, 113
- APT: Assessments of Performance in Teaching, 113
- CCI: Connecticut (teacher) Competency Instrument, 113
- FPMS: Florida (teacher) Performance Measurement System, 110, 113
- GTEP: Georgia Teacher Evaluation System, 113
- LCET: Louisiana Components of Effective Teaching, 113
- MTAI: MississippiTeacher Assessment Instruments, 113

TADS: Teacher Assessment and Development System (Dade Co., FL), 113
TAP: Teacher Assessment Process (Pasco Co., FL), 113
TCLTES: Tennessee Career Ladder Teacher Evaluation System, 113
TPAI: Teacher Performance Assessment Instruments, 113
TTAS: Texas Teacher Appraisal System, 113
VTPR: Virginia Teacher Performance Record, 113
Teacher's Work Sample Methodology (TWSM), 174-175
  analysis, 175
  data collection, 175
  interpretation, 175
  reflection, 175
  reporting, 175
  teacher capability-building costs, 175
  teacher portfolio, 175
Teaching models, 34
  behavior modification, 34
  information processing, 34
  personal, 34
  social interaction, 34

CORWIN
PRESS

The Corwin Press logo—a raven striding across an open book—represents the happy union of courage and learning. We are a professional-level publisher of books and journals for K–12 educators, and we are committed to creating and providing resources that embody these qualities. Corwin's motto is "Success for All Learners."

Evaluating Teaching

A Guide to Current Thinking and Best Practice

Editor:
James H. Stronge
College of William & Mary, Williamsburg, Virginia

See a dramatic rise in your teachers' performance levels. Find out how to create a high-quality teacher assessment and evaluation system that will

- Provide a framework for continuing staff improvement
- Increase the value of your school's programs and services
- Enable your school to more easily accomplish its mission
- Improve students' learning levels

Design, implement, and monitor an effective teacher assessment and evaluation system for your school. Use this coherent, comprehensive guidebook to make sure your teacher evaluation system is relevant, mutually beneficial, and participatory.

Find out what you can do to obtain an accurate picture of your staff's performance. Establish mutual goals and emphasize open communication for your teacher evaluations. Gather your data from multiple sources and use it to create an assessment environment that fosters trust.

Employ direct observation, "client" feedback, student performance data, teacher self-evaluation, and teacher portfolios to enhance the validity of your analyses. You'll be able to obtain a more detailed, accurate, and fair portrait of your teachers.

You'll find numerous illustrations and examples throughout the book. At the end of each chapter, questions and answers help you translate theory into practical use.

Your staff needs an effective teacher evaluation system to acomplish its best work. Make sure your system gives teachers the assistance they need to measure up to the highest standard of good teaching: successful student learning.

ISBN 0-8039-6356-4
9 780803 963566
90000>

CORWIN PRESS, INC.
A Sage Publications Company
2455 Teller Road
Thousand Oaks, CA 91320-2218
E-mail: order@corwin.sagepub.com
Call: 805-499-9774 Fax: 805-499-0871

BEST COPY AVAILABLE